Real Time Structured Light and Applications

Wilm, Jakob

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Real Time Structured Light and Applications

Jakob Wilm

Kongens Lyngby 2015

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

Structured light scanning is a versatile method for 3D shape acquisition. While much faster than most competing measurement techniques, most high-end structured light scans still take in the order of seconds to complete.

Low-cost sensors such as Microsoft Kinect and time of flight cameras have made 3D sensor ubiquitous and have resulted in a vast amount of new applications and methods. However, such low-cost sensors are generally limited in their accuracy and precision, making them unsuitable for e.g. accurate tracking and pose estimation.

With recent improvements in projector technology, increased processing power, and methods presented in this thesis, it is possible to perform structured light scans in real time with 20 depth measurements per second. This offers new opportunities for studying dynamic scenes, quality control, human-computer interaction and more.

This thesis discusses several aspects of real time structured light systems and presents contributions within calibration, scene coding and motion correction aspects. The problem of reliable and fast calibration of such systems is addressed with a novel calibration scheme utilising radial basis functions [Contribution B]. A high performance flexible open source software toolkit is presented [Contribution C], which makes real time scanning possible on commodity hardware. Further, an approach is presented to correct for motion artifacts in dynamic scenes [Contribution E].

An application for such systems is presented with a head tracking approach for medical motion correction [Contribution A, F]. This aims to solve the important problem of motion artifacts, which occur due to head movement during long acquisition times in MRI and PET scans. In contrast to existing methods, the one presented here is MRI compatible [Contribution D], not dependent on fiducial markers, and suitable for prospective correction.

Factors contributing to accuracy and precision of structured light systems are

investigated with a study of performance factors [Contribution G]. This is also done in the context of biological tissue, which exhibit subsurface effects and other undesirable effects [Contribution H], and it is shown that this error is to a large extent deterministic and can be corrected.

Summary (Danish)

Struktureret lys scanning er en alsidig metode til 3D måling, som blandt andet er meget hurtig. Dog tager de fleste målinger stadig i størrelsesordenen sekunder at gennemføre. Microsoft Kinect og "time of flight"-kameraer har gjort 3D sensorer allestedsnærværende og resulteret i et stort antal nye anvendelser og metoder. Disse sensorer er imidlertid begrænsede i deres nøjagtighed og præcision, hvilket gør dem uegnede til fx nøjagtig "tracking" og "pose estimation".

Med den nyeste udvikling inden for kameraer og projektorer, hurtigere regnekraft og metoder præsenteret i denne afhandling, er det muligt at udføre struktureret lys-skanninger i real-tid med 20 dybdemålinger per sekund og mere. Dette giver helt nye muligheder for studier af dynamiske fænomener, kvalitetskontrol, menneske-computer interaktion og mere.

Afhandlingen omhandler en række aspekter vedrørende real-tids struktureret lys, heriblandt kalibrering, spatial kodning og bevægelseskorrektion. Kalibreringsrutiner ønskes hurtige og robuste, hvilket addresseres med en ny kalibreringsmetode, som gør brug af radiale basis-funktioner [Bidrag B]. Et fleksibelt open source software program bliver præsenteret [Bidrag C], som muliggør real-tids scanning på almindelig hardware. Desuden præsenteres en metode til korrektion af bevægelsesartefakter, som optræder i dynamiske skanninger [Bidrag E].

En anvendelse af denne teknik bliver præsenteret med et system til hovedtracking til medicinsk bevægelseskorrektion [Bidrag A, F]. Dette har til formål at forbedre kvaliteten i skanninger med bevægelsesartefakter, som optræder ved lange optagetider i MR og PET scanninger. Modsat eksisterende metoder, er denne MR-kompatibel [Bidrag D], ikke afhængig af markører, og anvendlig til prospektiv korrektion.

Faktorer som har indflydelse på nøjagtigheden og præcisionen af struktureret lys systemer bliver undersøgt i et studie [Bidrag G]. Dette gøres også specifikt for biologisk væv, som bl.a. udviser "subsurface"-effekter [Bidrag H], og det bliver vist, at fejlen har en stor deterministik komponent, som kan korrigeres.

Preface

This work was done at the Section for Image Analysis and Computer Graphics at the Department for Applied Mathematics and Computer Science at the Technical University of Denmark (DTU Compute). The work is done in accordance with the programme of the department's PhD School (ITMAN) for acquiring the PhD degree.

Some parts of this work were done in collaboration with and at the Department of Clinical Physiology, Nuclear Medicine and PET at the Copenhagen University Hospital, Rigshospitalet. Part of the work was also done at the Athinoula A. Martinos Center for Biomedical Imaging, Boston Massachusetts, during an external stay of 4 months.

The PhD was supervised during the first two years by Professor Rasmus Larsen. In late 2014, Rasmus Larsen became Head of Department, and the new principle supervisor was Associate Professor Henrik Aanæs. Co-supervisors were Professor Liselotte Højgaard, Copenhagen University Hospital and Associate Professor Rasmus R. Paulsen.

Lyngby, 14-December-2015

Acknowledgements

First and foremost I would like to thank my supervisors, Professor Rasmus Larsen, Associate Professor Henrik Aanæs, Associate Professor Rasmus R. Paulsen and Professor Liselotte Højgaard.

Rasmus Larsen was my thesis supervisor during the first two years, and provided much help and assistance with both theoretical discussions as well as practical considerations. He then became Head of Department for DTU Compute, taking his time from supervision duties.

I was fortunate to have Henrik Aanæs as my new supervisor. This led my thesis work into a slightly different direction, with more emphasis on structured light scanning in general. Henrik has been enormously supportive and very kind in taking the time to discuss all aspects of our work.

The great work and initiative of Post.Doc. Oline V. Olesen has initiated this PhD project. I would like to thank her, as well as the clinical collaborators at Rigshospitalet, namely Professor Liselotte Højgaard. She was very kind to provide the facilities and funds to realise a prototype 3D scanner for medical motion correction.

I would also like to thank my good colleagues at the section for Image Analysis and Computer Graphics, DTU Compute. Especially the people with whom I share offices and supervisor, and have had many fruitful conversations and collaborations. These include Eyþór Rúnar Eiríksson, Jannik Boll Nielsen, Jonathan Dyssel Stets, Mads Doest, Sebastian Nesgaard Jensen, and Rasmus Ahrenkiel Lyngby.

Lastly, I would like to thank my girlfriend, Jenny, and my little son, Emil, for being very supportive and patient while I was working on my PhD.

Contributions

Thesis Contributions

The following lists the contributions which are part of this thesis in their order of publication. These consist of 5 peer-reviewed conference papers, one peer-reviewed conference poster, one technical report and one international patent. The contents of these contributions build the basis for the thesis. The full-text documents and short descriptions are found in appendices A – H, corresponding to the letters of the citation name.

[A] J. Wilm, O.V. Olesen, R.R. Jensen, L. Højgaard, and R. Larsen, *Fast and Practical Head Tracking in Brain Imaging with Time-of-Flight Camera*, IEEE Nuclear Science Symposium Conference Record, 2013. doi:10.1109/NSSMIC.2013.6829270

[B] J. Wilm, O.V. Olesen, and R. Larsen, *Accurate and Simple Calibration of DLP Projector Systems*, SPIE Photonics West, 2014. doi:10.1117/12.2038687.

[C] J. Wilm, O.V. Olesen, and R. Larsen, *SLStudio : Open-Source Framework for Real-Time Structured Light*, IEEE IPTA, 2014. doi:10.1109/IPTA.2014.7002001.

[D] O.V. Olesen, J. Wilm, A. van der Kouwe, R.R. Jensen, R. Larsen, and L.L. Wald, *An MRI Compatible Surface Scanner*, ISMRM Scientific Workshop, 2015. doi:10.1118/1.596566.

[E] J. Wilm, O.V. Olesen, and R.R. Paulsen, *Correction of Motion Artifacts for Real-Time Structured Light*, SCIA, Springer LNCS 9127 pp. 142–151, 2015. doi:10.1007/978-3-319-19665-7

[F] O.V. Olesen, R. Larsen, J. Wilm, R.R. Jensen, *Method for Surface Scanning in Medical Imaging and Related Apparatus*, International Patent, 2015. WO 2015/071369 A1.

[G] E.R. Eiríksson, J. Wilm, D.B. Pedersen, and H. Aanæs, *Precision and Accuracy Parameters in Structured Light 3-D Scanning*, ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016. doi:10.5194/isprsarchives-XL-5-W8-7-2016

[H] J. Wilm, S.N. Jensen, and H. Aanæs, *Structured Light Scanning of Skin, Muscle and Fat*, Technical Report, DTU Orbit, 2015.

Other Contributions

The following contributions were prepared and published by the thesis author in part during the time of the PhD, but are not directly related to the thesis. These consist of 4 peer-reviewed conference articles and 1 journal article.

- X. Cheng, J. Wilm, S. Seshamani, M. Fogtmann, C. Kroenke, and C. Studholme, *Adapting Parcellation Schemes to Study Fetal Brain Connectivity in Serial Imaging Studies*, IEEE EMBS, 2013. doi:10.1109/EMBC.2013.6609440
- M. Fogtmann, S. Seshamani, C. Kroenke, X. Cheng, T. Chapman, J. Wilm, C. Studholme, *A Unified Approach to Diffusion Direction Sensitive Slice Registration and 3-D DTI Reconstruction from Moving Fetal Brain Anatomy*, IEEE Transactions on Medical Imaging, 33(2) 2014. doi:10.1109/TMI.2013.2284014
- O.V. Olesen, J. Wilm, R.R. Paulsen, L. Højgaard, and R. Larsen, *DLP Technology Application: 3D Head Tracking and Motion Correction in Medical Brain Imaging*, SPIE Photonics West, 2014. doi:10.1117/12.2035374
- J.B. Nielsen, E.R. Eiríksson, R.L. Kristensen, J. Wilm, J.R. Frisvad, K. Conradsen, H. Aanæs, *Quality Assurance Based on Descriptive and Parsimonious Appearance Models*, MAM2015: Eurographics Workshop on Material Appearance Modeling. doi:10.2312/mam.20151199
- A. Dal Corso, M. Olsen, K.H. Steenstrup, J. Wilm, S.N. Jensen, R.R. Paulsen, E.R. Eiríksson, J.B. Nielsen, J.R. Frisvad, G. Einarsson, H.M. Kjer, *VirtualTable : a Projection Augmented Reality Game*, SIGGRAPH Asia 2015. doi:10.1145/2820926.2820950

Acronyms

CMM	Coordinate Measuring Machine
CT	Computed Tomography
DLP	Digital Light Processing
DMD	Digital Micromirror Device
EKF	Extended Kalman Filter
FOV	Field of View
FTP	Fourier Transform Profilometry
ICP	Iterative Closest Point
ISO	International Organization for Standardization
LCoS	Liquid Crystal on Silicone
LCD	Liquid Crystal Display
MEMS	Microelectromechanical System
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
PMD	Photon Mixer Device
PMP	Phase Modulation Profilometry
PS	Phase Shifting
RBF	Radial Basis Function
RF	Radio Frequency
RMSE	Root Mean Square Error
SL	Structured Light
SLS	Structured Light Scanning
SNR	Signal to Noise Ratio
TOF	Time of Flight

Contents

Chapter 1

Introduction

Active optical methods for dense measurements of object surfaces employ one or multiple light source and one or multiple cameras for shape acquisition. They have a wealth of uses and applications, and have been used in areas such as robotics, medicine, computer games, metrology and more. But only with recent progress in sensor technology and processing power, have they become accurate, fast and versatile enough to become widely used. Cornerstones of this development were the first SL systems [PA82], invention of the Photon Mixer Device (PMD) for Time of Flight (TOF) imaging [SXH+97], the introduction of the Microsoft Kinect 1 single-shot structured light system, and in recent years, the general availability of highly accurate systems (down to single digit micron) from manufacturers like GOM, Hexagon Metrology, and Aicon 3D Systems (see Fig. 1.1).

A number of distinct methods for active optical scene acquisition have been developed in recent decades. These fall into two categories: range-sensors, which rely on the finite speed of light, and triangulation based sensors which rely on view disparity. The first group includes TOF while the latter includes laser point and laser line scanning. Probably the most flexible group of methods is Structured Light Scanning (SLS), which is the main focus of this thesis. This technique has been available since the early 1980's, and its invention can be attributed to Posdamer and Altschuler [PA82]. A few alternative definitions of SLS are found in literature, e.g.[Gen11, SFPL10], with disagreement about the questions whether SLS uses exactly one or multiple cameras and whether single pattern coding qualifies. We define SLS as follows:

> SLS is the process of acquiring 2.5D depth images using an active light source, which projects one or more patterns, and one or more cameras, which capture images of these patterns projected onto the

Figure 1.1: Commercial optical measurement systems. Top left: PMD TOF camera. Top right: Microsoft Kinect 1. Bottom: GOM high-end structured light system.

scene. The encoded images are used to determine correspondence points, which are triangulated to yield 3D scene points.

With this definition, triangulation can occur between a camera and projector or between two or more cameras. In the latter case, the projected light can be considered a source of unique scene texture, which can also be categorised as "active stereo".

Structured light is a particularly attractive modality for optical shape acquisition for the following reasons:

- Flexibility: many applications require specific Field of View (FOV), accuracy/precision or illumination wavelength. Many of these requirements can be met by constructing the system accordingly, and by using a pattern coding method with the desired properties.
- Acquisition speed: in contrast to other active methods such as laser line scanning, structured light coding schemes have been developed which encode a scene with multiple thousands of scene points with as little as one static pattern. As an example, with binary coding the number of patterns and camera frames, N, is in the order of 10, and 2^N unique lines can be encoded. In laser line scanning, 2^N camera frames are needed for the same coverage.

- Precision: passive stereo methods generally suffer from low precision in textureless areas. Structured light adds texture to such areas, yielding precise point correspondences.
- Non-contact: in contrast to traditional high-accuracy measurement methods such as Coordinate Measuring Machine (CMM) in which a contact probe touches an item to digitise it.
- Low-cost: the components comprising an SL scanner are available as commodity hardware, and as such, highly accurate systems can be constructed at comparatively low prices.

Figure 1.2 shows the principle of structured light with a single projector-camera pair. In this particular case, correspondence points are encoded using sinusoidal patterns with the Phase Shifting (PS) method.

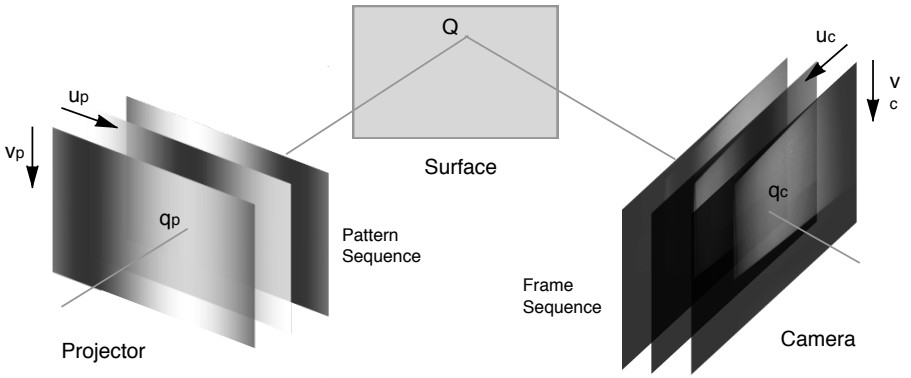


Figure 1.2: Principle of structured light with a single camera-projector pair. The projector projects a series of patterns (in this case a 3-step PS sequence), which are captured by the camera. This allows to identify the projector's u_p coordinate at that point, whose 3-D coordinates can then be determined. Modified from [Contribution C].

With this technique, a dense reconstruction of the surface can be obtained with as little as 3 projected patterns and camera images. This is in contrast to laser line scanning, where each camera image yields only reconstruction of a single line of points on the object surface [LRG+00]. However, the triangulation principle is the same, and as such, a SL scanner can be considered a parallelised laser scanner.

While SL has found use in many applications, particularly for measurement tasks requiring high accuracy, high speed or real time SL potentially enables a large

number of whole new applications from precise object tracking, the ability to study dynamic phenomena to human-computer-interaction, interactive games, biometric identification and more. A medical problem, which motivated this thesis work, is head motion during neurological Magnetic Resonance Imaging (MRI) scans [MHSZ13]. The approach presented here to alleviate this problem is by means of real time SL and accurate pose tracking.

Important challenges in SL, be it for accurate measurements or head tracking, are related to materials which have non-ideal optical properties. For this reason, the amount of error and the involved effects are also discussed in depth in this thesis.

1.1 Scope

The thesis has the following scope:

> The thesis revolves around structured light 3D scanning in real time. It shows how such data can be use for accurate rigid tracking. A practical use case is presented with head tracking and motion correction in medical scanners. Factors pertaining to the accuracy and precision of structured light scans are discussed in depth.

1.2 Objectives

Motivated by the potential of real-time SL, the following objectives are formulated for this thesis work:

1. To realise real time SL, capable of projecting, capturing, reconstructing and visualising dense reconstructions of a scene.
2. To perform real time pose tracking using SL.
3. To make such a system suitable for head motion correction in medical scanners.
4. To investigate the factors influencing measurement accuracy under ideal conditions.
5. To characterise and correct for measurement errors on biological tissue.

The following chapters will treat foundations relating to these objectives. The contributions in appendices A – H, in conjunction with these chapters, aim to fulfil the objectives.

1.3 Structure

This thesis acts as an overview and review of the current state-of-art within SL, real-time tracking, medical motion correction and the factors influencing SL performance. It should be self-contained, but also acts as an introduction and explanation of the contributions found in the appendix. The structure is as follows:

Ch. 2 introduces structured light methods on a conceptual level with the fundamental principles underlying it. This will provide the necessary foundation for real time SL and also for accuracy assessment and corrections as considered in later chapters.

In Ch. 3 a discussion of aspects relating to real time SL is provided with respect to hardware, pattern projection and computational efficiency. Implementation details are provided for a real time system.

Ch. 4 shows that the so-obtained data can be used for fast object tracking. The chapter also considers filtering of the tracking signal and its validation.

Ch. 5 then shows how such object tracking can be used for motion tracking and correction in MRI-scanners, with unique challenges and their solution.

The large amount of factors influencing accuracy and precision in SL are considered in Ch. 6.

A conclusion is lastly given in Ch. 7, which returns to the objectives mentioned above and points towards the future of this field.

Chapter 2

Structured Light Methods

With the definition of SLS on page 1, a discussion of the principles, theoretical aspects, and previous work surrounding this method shall be provided. While the field is extensive, and has seen many innovations over multiple decades, this discussion will focus on those aspects that relate to our objectives and contributions, with the hope of giving an overview of the field. Within this chapter, these are questions regarding calibration and coding strategies. These questions are important in any practical structured light system, and build the foundation for the development of real time SLS (Ch. 3) and also the applications presented in Ch. 4 and 5.

2.1 System Calibration

The most important factor for the accuracy that can be obtained with SL systems is a precise and accurate calibration of the camera, the projector (intrinsic parameters), and their spatial interrelationship – the extrinsic parameters. Formally, both devices are often modelled using the linear projective pin-hole model, augmented with a non-linear lens distortion model [HZ03, HS97]. The linear camera model is

$$\boldsymbol{P} = \boldsymbol{A}\,[\boldsymbol{R}|\boldsymbol{T}] = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [\boldsymbol{R}|\boldsymbol{T}] \quad , \tag{2.1}$$

in which $\boldsymbol{A}$ contains the so-called intrinsic parameters: focal lengths f_x, f_y and principal point coordinates c_x, c_y. More elaborate camera models exist [HZ03], however, with modern digital cameras of moderate focal length (non-fisheye), they are rarely necessary. The extrinsic parameters are given by rotation matrix

$\boldsymbol{R}$ and translation vector $\boldsymbol{T}$, and determine the camera's position and orientation in a world coordinate system. These are usually chosen such that one camera's position coincides with the origin, with it's view-direction along a principle coordinate axis.

The linear pin-hole model above is often extended by means of 5-parameter lens distortion parameters, $\boldsymbol{k} = [k_1, k_2, k_3, k_4, k_5]^\top$, as proposed by Brown [Bro71]. Lens barrel distortion is defined by three parameter in a radial function:

$$s(x, y) = 1 + k_1 r^2 + k_2 r^4 + k_5 r^6, \quad r^2 = x^2 + y^2 \quad , \tag{2.2}$$

while imperfect alignment of lens elements gives rise to "tangential distortion", which is modelled as

$$\boldsymbol{t}(x, y) = \begin{bmatrix} 2k_3 xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4 xy \end{bmatrix}, \quad r^2 = x^2 + y^2 \,. \tag{2.3}$$

These functions relate the ideal, distortion-free image coordinates (u, v) to the observed coordinates $(\tilde{u}, \tilde{v})$ in the following way:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} + s(x, y) \begin{bmatrix} x \\ y \end{bmatrix} + \boldsymbol{t}(x, y), \quad x = \frac{u - c_x}{f_x}, \quad y = \frac{v - c_y}{f_y} \quad . \tag{2.4}$$

2.1.1 Camera Calibration

Camera calibration is the process of determining the camera and lens model parameters accurately. With the pinhole model, this amounts to determining $\boldsymbol{A}$, $\boldsymbol{R}$, $\boldsymbol{T}$. It is usually sufficient to estimate $\boldsymbol{P}$, as it can be decomposed into these three matrices. Lens distortion parameters $\boldsymbol{k}$ are often estimated separately in a non-linear optimisation step.

In "auto-" or "self-calibration", parameters are determined from normal camera images viewing a general scene [FLM92, Har94]. Depending on the specific method, little or no assumptions are made about the viewed scene, or the motion of the camera between images.

The alternative is traditional offline calibration, in which specific calibration images are taken under specific constraints. The calibration object defines a world coordinate system such that 3D coordinates of the visual features are known. Most of these methods work by observing a calibration object with known visual features. This is preferred when full control over the calibration procedure is necessary and high accuracy is demanded. An early method is that of Tsai [Tsa86]. More intrinsic parameters were included by Heikkilä and Silvén [HS97].

A modern and popular method in the computer vision community is that of Zhang [Zha99], which also is implemented in popular software libraries [Bra00, Bou, Mat]. This calibration routine relies on observations of a planar calibration board with easily recognisable features. It relates the 3-D coordinates of these to the observed image coordinates projections by means of the model above, and solves for the calibration plane extrinsics (the camera's position and orientation relative to the calibration board's coordinate system), and the camera intrinsics, $\boldsymbol{A}$ by means of a closed form solution. This is then followed by non-linear optimisation with the Levenberg-Marquardt algorithm over all parameters, including $\boldsymbol{k}$. The objective function to be minimised is the sum of squared reprojection errors, defined in the image plane (from [Zha99]):

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\boldsymbol{p}_{ij}-\breve{\boldsymbol{p}}(\boldsymbol{P}_j,\boldsymbol{A},\boldsymbol{k},\boldsymbol{R}_i,\boldsymbol{T}_i)\right\|^2 \quad , \tag{2.5}$$

where $\breve{\boldsymbol{p}}$ is the projection operator determining 2-D point coordinates given 3-D coordinates and the camera parameters as defined in Eqn. 2.1. i sums over the positions of the calibration board and j over the points in a single position. $\boldsymbol{P}_j$ are 3-D point coordinates in the local calibration object coordinate system, $\boldsymbol{P}_j = [x, y, 0]^\top$, and $\boldsymbol{p}_{ij}$ the observed 2-D coordinates in the camera. The per-position extrinsics $\boldsymbol{R}_i, \boldsymbol{T}_i$ can be understood as the cameras position relative to coordinate system defined by the calibration object. With quality lenses and calibration targets, final mean reprojection errors in the order of 0.1 px are usually achieved.

Zhang's method is stable in many configurations, but works with planar 2-D calibration objects only. Such planar calibration objects with checkerboard or circular control point targets are very popular due to their simple construction. For instance, Fig. 2.1 shows a planar float glass calibration target with a checkerboard pattern. Saddle points may be detected and located automatically up to very high accuracy, and invariantly to the angle of observation.

2.1.2 Projector Calibration

In SL systems that utilise a single camera, the projector must also be calibrated. While camera-calibration is well studied, and can be performed up to very high accuracy, the main problem in single camera-projector SL systems is the determination of these projector parameters. Since a calibration object cannot be directly observed by the projector, different strategies are needed to infer the parameters of the projector.

One option is to not determine the projector parameters explicitly, but to de-

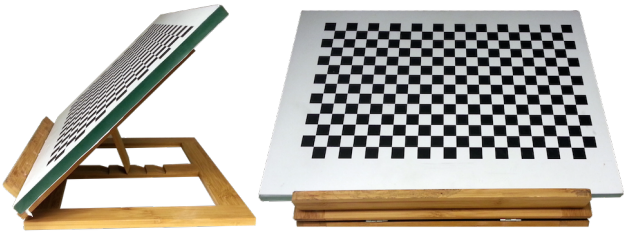


Figure 2.1: Planar calibration target used for camera calibration according to Zhang's method [Zha99]. The target is made of 10 mm unhardened float glass with a laser-printed checker pattern glued onto it. Float glass is temperature stable, very planar before hardening and readily available. From [Contribution G].

code a dense structured light sequence on a reference plane [ZS94]. Subsequent measurements are then relative to this reference. This also gives computational advantages in point triangulation, as discussed in Sec. 2.4.2. However, it does not provide the full projector parameterisation, which is useful in many algorithms. Additionally, it is based on noisy measurements at a certain distance, and cannot be assumed to be accurate elsewhere [MWZ13].

A number of methods for projector calibration require a calibration board to be positioned accurately e.g. by means of translation stages or in fronto-parallel orientation [GHYC04, CXJS09, ADP12, DRS09]. These methods are not convenient in many situations, where such positioning equipment is not available.

Some approaches use a flat planar board with a checkerboard pattern occupying part of the space. The board's pose is then determined using a calibrated camera. Feature points are projected onto it, and 3-D point coordinates determined by means of ray-plane intersection or through a homography [KMK07, SWPG05, LC08, GWH08, GWH08, OJP+10, PP10, SC08]. A disadvantage with these methods is that a global homography is linear, and as such cannot model the non-linear lens distortion. Additionally, it is difficult to span the entire camera and projector FOV with feature points. Lastly, it is difficult to automate feature point detection when multiple patterns are present.

The concept of indirectly "capturing" images is due to Zhang and Huang [ZH06] and adapted in a number of subsequent methods, e.g. [LSWW08, YPL+12]. A similar method was also presented by [LS04], albeit without much detail. In this concept, the projector encodes its coordinates onto a calibration object by means of a SL sequence, which is then captured by the camera. The method of indirectly capturing images using the projector allows one to establish a map-

ping from camera to projector space, providing the homologous feature point coordinates in the projector's coordinate system. These can then be used as the input to regular camera calibration for the projector, and stereo calibration for the system. See Fig. 2.2 for a flow diagram of this approach. This concept has the advantage that the error of camera calibration does not propagate into projector parameter estimates, making it potentially very accurate. The disadvantage is that a sequence must be projected for each position and orientation of the calibration object. In addition, coordinate conversion must be performed robustly and accurately. To address this potential issue, Moreno and Taubin define "local homographies" [MT12], which provide a linear relationship between camera and projector point coordinates in the vicinity of a feature point.

This thesis' [Contribution B] provides a very accurate method for calibration also based on Zhang-Huang's method. It uses Radial Basis Functions (RBFs) to locally translate camera to projector coordinates and estimate the feature point locations in projector space with sub-pixel precision. This is both robust and accurate in that the non-linearity of this translation can be accommodated with RBFs.

2.1.3 Stereo Calibration

Stereo calibration is the process of determining relative extrinsic parameters between two rigidly coupled cameras. These can be parameterised by means of rotation matrix $\boldsymbol{R}_\mathrm{s}$ and translation vector $\boldsymbol{T}_\mathrm{s}$. This is relevant in both single-camera SL systems, and those based on two cameras. With traditional offline calibration (as opposed to self-calibration), the process is usually very similar to camera calibration in that a linear initialiser is first computed, followed by non-linear optimisation using e.g. the Levenberg-Marquardt algorithm with the total reprojection error as the objective function. Both camera's intrinsic parameters can be adjusted to minimise the stereo reprojection error, but often it is beneficial to perform single camera calibrations and only estimate $\boldsymbol{R}_\mathrm{s}$ and $\boldsymbol{T}_\mathrm{s}$ in stereo calibration [MT12]. The stereo reprojection error to be minimised is

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \left\|\boldsymbol{p}_{1,ij} - \breve{\boldsymbol{p}}(P_j, \boldsymbol{A}_1, \boldsymbol{k}_1, \boldsymbol{R}_i, \boldsymbol{T}_i)\right\|^2 + \left\|\boldsymbol{p}_{2,ij} - \breve{\boldsymbol{p}}(P_j, \boldsymbol{A}_2, \boldsymbol{k}_2, \boldsymbol{R}_\mathrm{s}\boldsymbol{R}_i, \boldsymbol{R}_\mathrm{s}\boldsymbol{T}_i + \boldsymbol{T}_\mathrm{s})\right\|^2 \quad , \tag{2.6}$$

where $\breve{\boldsymbol{p}}$ is the projection operator, $\boldsymbol{p}_{1,ij}$ the point coordinates as seen in the first camera, and $\boldsymbol{p}_{2,ij}$ are those seen in the other. i sums over the positions of the calibration board and j over the points in a single position. $P_j = [x, y, 0]^\top$ are 3-D point coordinates in the local checkerboard coordinate system. With quality lenses and calibration targets, final stereo mean reprojection errors in the order of 0.1 px are often achieved.

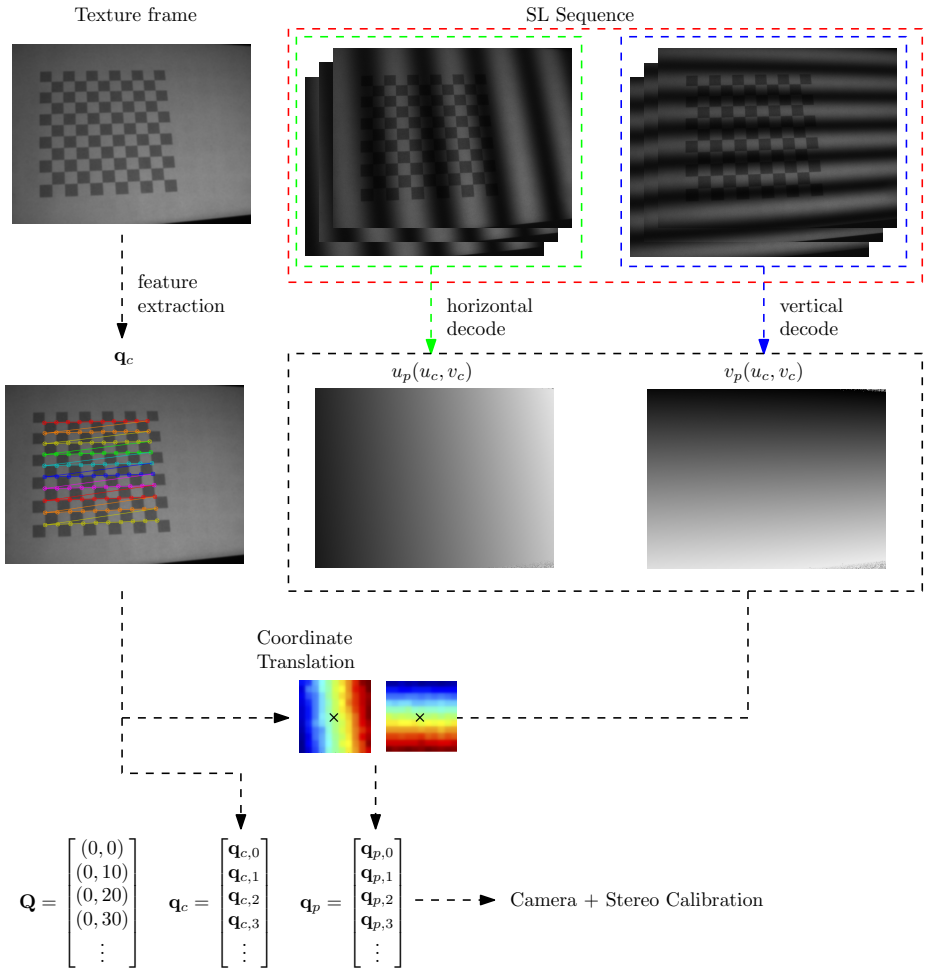


Figure 2.2: Flow diagram of the projector calibration technique according to Zhang and Huang [ZH06]. An SL sequence is projected onto the calibration object – in this case a planar chessboard pattern target. The resulting camera and projector coordinates are used to translate the sub-pixel feature point coordinates into projector space. The resulting points coordinates are used in camera and stereo calibration. Modified from [Contribution B].

2.1.4 Lens Correction

As mentioned, the lens distortion and non-linearity is usually considered to have a radial and tangential component (see Eqn. 2.2 and Eqn. 2.3). With the calibration method of [Contribution B], these parameters can accurately be estimated for both camera and projector. In order to incorporate the camera lens distortion in the reconstruction process and improve the accuracy of point clouds, the distortion model must be applied inversely to the camera frames or to the reconstructed points, which can be more accurate as it does not involve interpolation in images.

Correcting for projector lens distortion is different in that the forward distortion model must be applied to patterns before projection. For both kinds of lens correction, warp fields can be pre-computed, which saves computing resources considerably, as the same correction needs to be applied many times.

2.2 Coding strategies

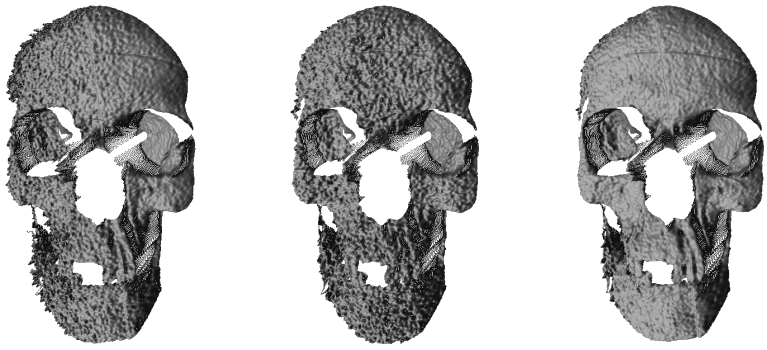
From a computer vision perspective, the most interesting question regarding SLS is: ”which pattern shall be used for scene coding?”. A large number of coding strategies have been suggested. Fig. 2.3 shows a representative selection of results obtained with different such methods.

2.2.1 Literature Overview

Since early research in structured light methods in the 1980's the community has proposed a wealth of coding strategies. Recent general reviews of structured light methods are found in [Gen11] and [SFPL10]. A review of *Phase Shifting* (PS) based techniques is given in [GR10]. An overview of the broader class of 3D optical methods is given in [CBS00].

Binary coding for structured light was proposed by Posdamer and Altschuler [PA82] as the earliest method for scene recovery that can be classified as ”structured light” according to the definition on page 1.

A different class of methods has its origin in analog optical methods such as Moiré Tophography [MJA70], Phase Modulation Profilometry (PMP) [SLH84] and Fourier Transform Profilometry (FTP) [TIK82, TM83, SC01]. PMP was performed using optical interferometers to generate sinusoidal light projections



(a) 2+1 Phase Shifting (3 patterns)

(b) 3-shift Phase Shifting (3 patterns)

(c) 4-shift Phase Shifting (4 patterns)

(d) Gray Coding (11 patterns)

(e) 3-step Phase Shifting (6 patterns) 16 Periods

(f) Micro Phase Shifting (12 patterns)

Figure 2.3: Comparison of various structured light coding strategies on a plastic skull with beneficial optical properties. Note that the number of patterns is not the same, and that a small amount of spatial smoothing was applied for the 2+1, 3-shift and 4-shift methods. Parameters were chosen to obtain the best possible visual results.

onto the measurement surface, which is both time-consuming and often imprecise. With digital video projectors, it has become significantly easier to generate accurate fringe projection patterns, giving rise to the term "digital fringe projection". An overview of fringe based techniques is given in [GR10].

Recent developments which have attracted some attention include "Modulated Phase Shifting" [CSL08], "Micro Phase Shifting"[GN12], "Ensembles of Codes"[GAVN12], and "Unstructured Light" [CMR11, CMR14]. These mainly deal with handling non-ideal optical properties of the scanned surfaces.

2.3 Phase Shifting Methods

Among the various coding strategies, PS based methods are particularly versatile, and suitable for a wide range of applications. Their strength is the ability to encode a scene completely with as little as three patterns (three phase steps), and scaling arbitrarily to N steps providing increasingly more precision. The essential phase shifting method encodes the scene close to along the epipolar direction (as will be discussed in Sec. 2.4), which here is assumed to be the horizontal u_p direction with the following N images:

$$I_n^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\left(\frac{n}{N} - \frac{u_p}{N_p}\right)\right) \quad , \tag{2.7}$$

where n indicates the pattern index, $n \in 1 \dots N$, and N_p is the number of projector columns. The intensity of the n'th pattern as observed in the camera can be expressed as

$$I_n^c(u_c, v_c) = A(u_c, v_c)^c + B(u_c, v_c)^c \cos\left(2\pi\left(\frac{n}{N} - \frac{u_p}{N_p}\right)\right) \quad , \tag{2.8}$$

in which $\frac{v_p}{N_p}$ denotes the corresponding phase of the sinusoidal sequence at that particular camera pixel. A^c is the intensity of the scene including ambient contributions, and B^c the intensity of reflected projector light. In terms of Fourier analysis, A^c can be considered the magnitude of the constant component, while B^c is the magnitude at the principle frequency, and θ its phase. As such, θ can easily be extracted by means of the fast Fourier transform and scaled to the number of projector columns. Figure 2.4 illustrates this principle. Figure 2.3b, 2.3c, and 2.3e show examples of surfaces reconstructed by means of PS-based methods.

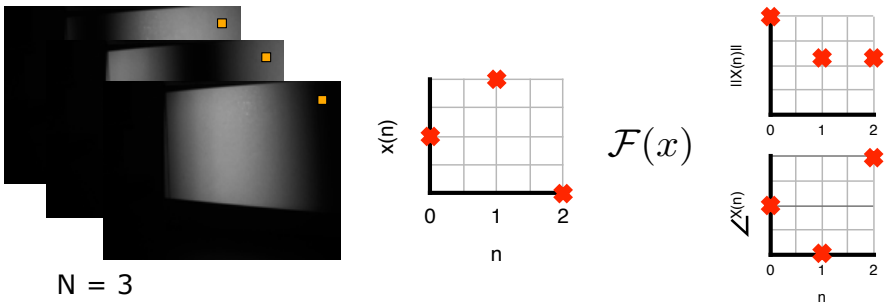


Figure 2.4: The principle in the PS scene coding technique with a single period pattern and $N = 3$ shifts. The correspondence information lies in the phase of the sequence observed at a specific pixel position. The corresponding amplitude provides a texture value.

The phase of the sinusoid at the principle frequency is given by:

$$\theta = \tan^{-1}\left(\frac{\sum_{n=1}^{N} I_n^c \cos(2\pi(n-1)/N)}{\sum_{n=1}^{N} I_n^c \sin(2\pi(n-1)/N)}\right) \quad , \tag{2.9}$$

which in the minimum case of the 3-step algorithm ($N = 3$) yields

$$\theta = \tan^{-1}\left(\frac{2I_1^c - I_2^c - I_3^c}{\sqrt{3}(I_2^c - I_3^c)}\right) \quad . \tag{2.10}$$

In most cases, a low number of phase steps is used, and multiple phases of the sinusoidal pattern are projected, e.g. [WZO11]. Using multiple phases leads to the so-called "phase wrapping" ambiguity, which needs to be resolved [ZLY07].

2.3.1 Phase Unwrapping

The wrapped phase problem occurs in a number of technical problems, e.g. magnetic resonance imaging and radar signalling. In phase shifting based SL, it is the consequence of using multiple phases in the projected sinusoid, which leads to ambiguity in the matching process [QTCF03, ZLY07]. The relationship between a 2 dimensional unwrapped phase and its unwrapped counterpart is:

$$\phi_{\text{unwrapped}}(x, y) = \phi_{\text{wrapped}}(x, y) + k(x, y)2\pi \quad k \in \mathbb{N} \quad . \tag{2.11}$$

This process is illustrated in Fig. 2.5.

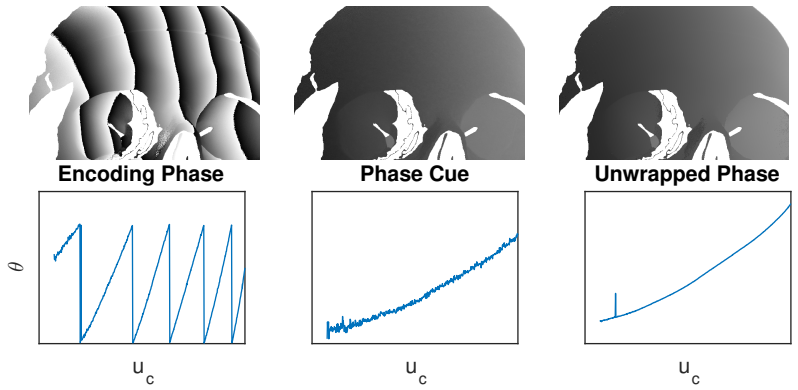


Figure 2.5: The principle of phase unwrapping by means of a phase cue. The bottom row plots show the phase along a single image row. Left: the multi-period coding phase is accurate but has $2-\pi$ ambiguities. Middle: a phase cue is created using a single-period sequence. Right: both results are combined to yield an accurate phase using Eq.2.11, however, unwrapping errors can occur in noisy regions (seen as a spike).

Unwrapping methods are either spatial and based on the raw multi-phase data or the unwrapping process is facilitated e.g. by multiplexing the signal with a special carrier. Alternatively, one can use additional patterns that disambiguate among the multiple phases – so-called temporal unwrapping which was originally proposed by Huntley and Saldner [HS93].

Temporal unwrapping is generally considered robust, and can be realised in different ways. In it's simplest form, an additional single-phase sequence is projected, which encodes every projector uniquely, albeit with much higher noise than the multi-phase patterns. The so-obtained "phase cue" is then used to add multiples of 2π to the multi-phase information and make it continuous. One problem with this strategy is that a low-frequency signal is used for the phase cue, while the main coding signal is of higher frequency. Due to global light effects and camera defocus, signals of different spatial frequency are altered differently, especially on surfaces which exhibit sub-surface scattering effects or reflections [NG12]. A detailed discussion of these effects will be given in Sec. 6.4.

For these reasons, temporal phase unwrapping may also be done with a second set of high-frequency patterns, which are chosen exactly such that their wavelength ratio satisfies the heterodyne principle [RRT97]. According to this principle, two wrapped phases of wavelength λ_1 and λ_2 may be subtracted from each other to yield a phase with the so-called beat-wavelength λ_b. This is illus-

trated in Fig. 2.6. λ_1, λ_2 and λ_b satisfy

$$\lambda_b = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \quad , \tag{2.12}$$

which allows one to choose wavelengths such that the beat wavelength equals the number of projector columns (or rows in the case of horizontal coding). The so determined beat phase can be used to unwrap the first or the second wrapped phase.

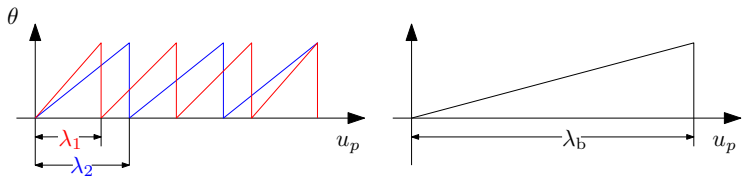


Figure 2.6: Illustration of the beat frequency principle. The sinusoids of wavelength λ_1 and λ_2 can subtracted to yield an equivalent signal of wavelength λ_b which can be used for unwrapping. Modified from [RRT97].

Both phase cue based unwrapping and unwrapping according to the heterodyne principle can naturally be performed at multiple levels, which can reliably unwrap high frequency patterns [RRT97].

In the "Micro Phase Shifting" method [GN12], a number of sinusoidal patterns are projected which belong to a narrow frequency band. This has the aforementioned advantage that both the primary frequency signal and the additional unwrapping patterns are affected similarly by various optical and material effects. In addition, "Micro Phase Shifting" uses an optimal number of patterns, which generally can be lower than with multi-frequency heterodyne patterns, and chosen to obtain a desired trade-off between phase unwrapping robustness, Signal to Noise Ratio (SNR), and accuracy.

2.4 Point Triangulation

Point triangulation amounts to the task of determining 3-D coordinates of points on the scanned object surface. Mathematically, point triangulation consists of solving for the 3D coordinates, $Q = [x, y, z, 1]^\top$, using the camera's projection matrix, $\boldsymbol{P}_c$, and the projector's projection matrix, $\boldsymbol{P}_p$. The projections of Q into camera and projector pixel-space are given by $\boldsymbol{q}_c = [u_c, v_c, 1]^\top$ and

$\boldsymbol{q}_p = [u_p, v_p, 1]^\top$. In the most general case, both horizontal and vertical directions are encoded in a SL sequence, and u_p and v_p are known at every camera pixel location (u_c, v_c). The estimation problem is then overdetermined (3 unknowns and 4 constraints), and a number of different triangulation algorithms may be used. In the simplest case, the "direct linear transform" methodology [HZ03] is applied to solve the overdetermined problem. This approach is fast but purely algebraic, and hence the minimised error not geometrically meaningful. In the "mid-point algorithm", Q is found as the midpoint of a line-segment connecting the viewing rays at their closest points. These methods are detailed e.g. in [HZ03].

2.4.1 Algebraic Triangulation

While scene coding in both horizontal and vertical directions is possible, in most SL systems, coding is done in only one direction [Gen11]. A requirement is that the coding method assigns unique values to scene points along the epipolar direction. In order for to this provide most useful information, the angle between coding direction and epipolar lines should be as small as possible, see Fig. 2.7.

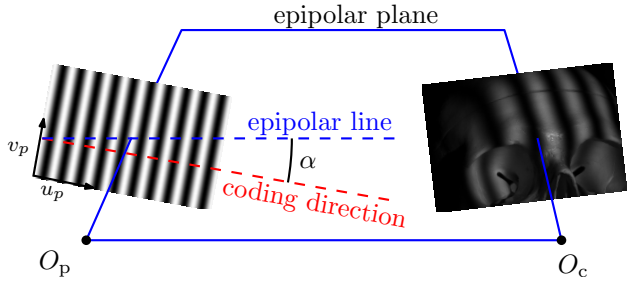


Figure 2.7: Relation of the coding direction to the epipolar lines as seen in the projector. The scene must be uniquely encoded along epipolar lines. For practical reasons the coding direction is along the principle axis (e.g. u_p). Angle α determines the degree to which the coding information constraints the triangulation problem, and should ideally be 0.

For simplicity, a horizontal layout as shown in Fig. 2.7 shall be assumed in the following. The projection of scene point Q into camera and projector space yields

$$\begin{aligned} \boldsymbol{P}_c Q &\propto \boldsymbol{q}_c \\ \boldsymbol{P}_p Q &\propto \boldsymbol{q}_p \end{aligned} \quad , \tag{2.13}$$

where $\propto$ denotes proportionality or homogenous equivalence. However, in this case only the the projector's column coordinate u_p was encoded. Thus, Q is recovered from the following homogenous equation:

$$\begin{bmatrix} \boldsymbol{P}_c(1) - u_c\boldsymbol{P}_c(3) \\ \boldsymbol{P}_c(2) - v_c\boldsymbol{P}_c(3) \\ \boldsymbol{P}_p(1) - u_p\boldsymbol{P}_p(3) \end{bmatrix} Q = 0 \quad . \tag{2.14}$$

$\boldsymbol{P}_x(i)$ indicates the i-th row of the matrix. Triangulation is computationally expensive, in that a linear system must be solved in every single camera pixel, where resolutions of a few million pixel are not uncommon. As will be shown in Sec. 3.4, however, solving this equation can be sped up considerably by pre-computing part of the solution.

2.4.2 Plane-Based Triangulation

With PS-based coding methods, one can potentially forego the algebraic triangulation altogether. The idea is to capture a structured light patterns on a planar object. The resulting phase image, ϕ, serves as a reference. The relative height to this plane can now be determined efficiently by means of simple trigonometry.

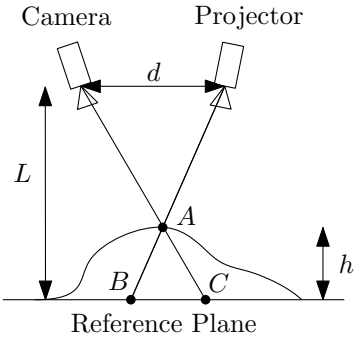


Figure 2.8: The geometry in plane-based reconstruction with PS based methods. Modified from [ZS94].

Referring to Fig. 2.8, one is interested in the distance of point A to the reference plane. The distance d between camera and projector centres, the distance to the reference plane, L, are determined during calibration. The height h from reference plane to point A can be found by considering similar triangles [ZS94]:

$$\frac{h}{|BC|} = \frac{L-h}{d} \Leftrightarrow h = \frac{|BC| \cdot L}{d + |BC|} \quad , \tag{2.15}$$

where $|BC|$ is directly proportional to the difference of measured phase at points B and C. With this simple relation, the scene may be reconstructed very efficiently. This has disadvantages though, in that the reference plane must be measured for each calibration of the SLS and the phase error from reference plane determination propagates into the final point measurements.

2.5 Contributions

SL is a well established technique with great benefits over other 3-D methods. One often arising issue is the accurate calibration of projectors, which is necessary in single projector-camera systems. A very accurate and fast method for SL calibration was developed and is presented in [Contribution B]. It is based on the observation of a planar calibration target which is illuminated with a SL sequence in multiple positions as shown in Fig. 2.2. It uses a novel RBF interpolation scheme to translate feature coordinates.

As part of the thesis work, a comprehensive open source SL software for single camera-projector SL was developed, and is presented in [Contribution C]. The calibration method of [Contribution B] is also implemented in this software.

In relation to calibration, a study of accuracy and precision parameters for SL, covering among other things calibration parameters and observations in a typical SL system was conducted and is included in [Contribution G].

Chapter 3

Real Time Structured Light

With most commercially available structured light systems and also with most setups described in literature, acquisition and reconstruction time is not of major concern. Hence, speed is often traded for accuracy with a large number of patterns in the coding sequence. As was argued in Ch. 1, with real time systems, a plethora of new applications become available, which is the main motivation of this thesis. Real time performance may be defined as providing dense point clouds at refresh rates comparable to what humans are able to perceive. A formal definition for these purposes is:

> Real time structured light is the projection, acquisition, processing and dense reconstruction of scenes with at least 10 point clouds per second containing at least 10.000 points and with latencies below 100 ms.

There are multiple challenges with real time systems. A short discussion of the main difficulties in the implementation of such systems shall be given in this chapter.

Projection speed in digital consumer projectors is usually designed for 60 Hz RGB video input. In most consumer products, various image processing (chroma sub-sampling, sharpening, colour conversion, gamma-correction) reduce image quality and increase latency.

Capture speed is a challenge as most industrial cameras have frame rates in the order of 30 Hz. Most cameras expose, read-out and transmit data sequentially, which sets additional requirements for projector-camera synchronisation. A carefully designed hardware synchronisation is thus necessary if multiple patterns are used in the SL sequence. With short exposure times, the projected light

output must also be correspondingly brighter. This does not always harmonise with human eye safety levels [Ole11].

The length of the SL pattern sequence is inversely proportional with the effective refresh rate. This favours e.g. single-shot techniques, or 3-step phase shifting with image-based unwrapping. These techniques also differ in decoding, which is usually processing intensive, and not all coding strategies are suitable for this reason.

Decoding and triangulation are other computationally expensive tasks, which affect overall performance and latency. For instance, in PS-based methods, the inverse tangent needs to be computed (Eqn. 2.9), and triangulation performed (solving Eqn. 2.14) at each pixel.

Thus, the main challenges with real time SL can be summarised as follows:

- Pattern projection speed and quality, limited mainly by the speed of pattern rendering and the projection hardware.
- Camera capture speed and synchronisation. The camera exposure must be precisely controlled and synchronised with the projection of patterns.
- Permitted light output for eye-safe operation, such that camera images are properly exposed. It is beneficial to use cameras with high sensitivity at the illumination wavelength.
- Sequence length, which is inversely proportional to the point cloud refresh rate obtainable.
- Computational requirements for decoding and triangulation.

This chapter introduces hardware and algorithms to overcome these challenges and presents the contributions made to the field.

3.1 Literature Overview

Relatively few implementations of real time structured light which satisfy the definition on page 23 have so far been described in literature.

Early real time systems mainly use one-shot strategies. The ”Rainbow 3D Camera” of Geng [Gen96] uses illumination of spatially varying wavelength. Wust

and Capson [WC91] and similarly Huang [HHJC99] utilise the three colour channels to project a 3-step PS pattern sequence onto the scene in a single projection. Zhang et al. use a single-shot colour de-Bruin pattern [ZCS02]. An elaborate technique with scene adaptive colour single-shot patterns is shown by Koninckx and Gool [KV06].

An early example of a system capable of real time reconstruction with multiple patterns is that Rusinkiewicz et al. [RHHL02], which uses multiple binary patterns, and tracks stripe boundaries as a means of motion correction. An early multi-pattern PS-based system is demonstrated by Huang et al. [HZC03]. They use a hardware modified consumer projector with the colour wheel removed. The accurate projection of patterns is tackled with a dedicated, analogue pattern projector of Wissmann [WSF11, WFS11]. The implementation of Zuo et al. [ZCG$^+$12] uses a modified commercial projector. Another real time PS-based system with an analytical method for error compensation is presented by Weise [WLV07]. In [ZXYW14], a PS/single-shot hybrid is presented which is able to switch between high fidelity and fast reconstruction depending on scene motion. A review of high speed PS-based methods and implementations is given in [SZ10].

The works of Lau and Hassebrook, University of Kentucky are influential. Their pattern compositing technique allows for simultaneous projection and subsequent separation of sinusoidal patterns in a one-shot pattern [GHL03] and two-frequency modulation halving the number of projected patterns in PS [LWL$^+$10a]. Other contributions include gamma correction [LWL$^+$10b] and noise modelling for PS frequency selection [LHG03, WLL$^+$10].

A large amount of innovation is due to the "Multi-scale Superfast 3D Optical Sensing Laboratory" of Zhang at Purdue University. His PhD thesis describes the implementation of 40 Hz real time structured light using consumer hardware [Zha05] and the group has presented various improvements to processing speed, albeit with quality trade-offs [Zha10], i.e. the use of deliberately defocused high-speed binary patterns with dithering [LZ09]. In [WLEZ13], a high speed camera is used to capture a beating rabbit-heart at 1000 Hz. A particularly fast image-driven phase-unwrapping technique was also developed [ZLY07]. Further improvements include GPU processing [KHCZ14]. The focus of this work is high speed, and few requirements are set for accuracy or permitted light intensity.

As for commercial systems, the popular Microsoft Kinect 1 sensor is an example of a single-shot structured light systems. It provides 30 depth and colour images per second. The single-shot pattern is a pseudo-random dot matrix projected from an infrared laser-projector. The main innovation of this sensor is a specialised integrated circuit developed by PrimeSense, which performs feature matching and triangulation inside the sensor with minimal power requirements.

The infrared camera has a resolution of 640×480 px, and the standard deviation of depth measurements is a few millimetres at 50 cm distance [KE12].

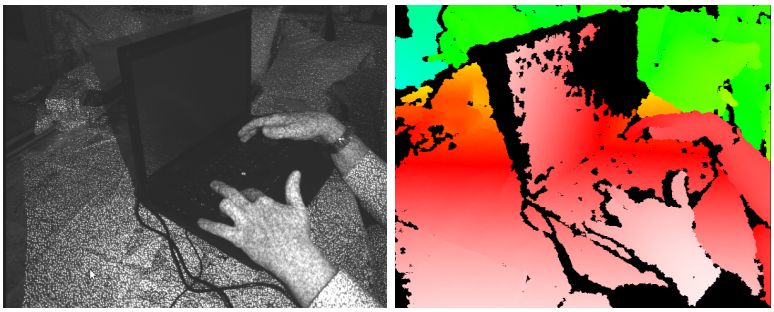


Figure 3.1: Left: infrared camera image of Kinect 1 showing the pseudo-random dot matrix. Right: the corresponding depth map. Black regions denote missing data due to low contrast or occlusion. (c) Wikimedia User:Kolossos.

3.2 Pattern Projection

In order to obtain high speed while maintaining the desirable properties of PS methods, such as high precision and accuracy, and robustness towards scene discontinuities and varying surface colour, true gray-scale patterns must be projected onto the scene. This can be realised by means of specialised projection hardware, such as that of Wissmann [WSF11]. Using a generic digital projection unit provides more flexibility however. Three projection technologies are currently available in consumer hardware: Digital Light Processing (DLP), Liquid Crystal on Silicone (LCoS) and Liquid Crystal Display (LCD) [FH09]. DLP is most widely used as it provides very high contrast and allows for high luminous output. It is based on a Microelectromechanical System (MEMS) called the Digital Micromirror Device (DMD). This unit consists of a large number of small mirrors which can be actuated electro-magnetically into one of two states, effectively projecting light out of the projector lens, or onto a heat-sink. With all these projection technologies, grey value projection relies on the integration of many binary patterns in the human eye or camera. This sets constraints on the allowable camera exposure time, as it generally needs to be a multiple of one projected frame period.

Currently available systems can achieve up to $9,523$ Hz binary or 247 Hz 8-bit grayscale patterns (December 2015, Texas Instruments LightCrafter 9000).

With a 3-step PS pattern sequence, this allows for potentially $82\frac{1}{3}$ surface reconstructions per second. However, to project patterns at these frequencies, they must generally be stored in fast access memory close to the projector controller hardware. A more generic approach, which allows for dynamic change of the sequence is to render the patterns on a computer's graphics unit, and transmit them via a display interface.

3.2.1 OpenGL Based Rendering

One of the advantages of rendering the SL sequence on the fly as opposed to storing it in hardware is the ability to adaptively adjust patterns based on scene change or complexity. Efficient rendering of structured light patterns requires direct access of the computer's graphics unit. The OpenGL library is a popular way of performing such rendering tasks with small overhead. As will be discussed in Sec. 3.3, any delays in the rendering process can lead to loss of synchronisation. Patterns can be efficiently rendered into offscreen framebuffers, which are then blitted on the graphics unit under operation. Fig. 3.2 shows the steps involved in this rendering solution.

A practical aspect of performing such pattern rendering is that the computer's interface should not interfere with the projection of patterns. This can be realised with the X screen mechanism on the Linux operating system by configuring two individual X screens. This also allows for vertical synchronisation to the projector, avoiding screen tearing. This thesis' [Contribution C] details an open source software platform which implements this rendering framework.

3.3 Synchronisation

An important practical aspect of SL is the correct capture of projected patterns. In low speed applications this can be achieved by synchronising the projection and capture loop in software. In this case, patterns are rendered and projected by the projector. The software sleeps to accommodate latency period ω, which is the time between rendering of the pattern until it is actually projected. This latency is due to double buffered rendering, transmission of the video signal, ring buffering in the projector and image processing performed by the projector, e.g. chroma sub-sampling, sharpening, colour conversion, gamma-correction (Texas Instruments DLPC350 documentation DLPS029B). This latency is usually in the order of $\omega = 50\,\text{ms}$. The camera is then software-triggered, which incurs a latency δ before image exposure begins. See Fig. 3.3 for a diagram of the timing

Pattern computation: $I_n^p(u_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\left(\frac{n}{3} - \frac{u_p}{N_p}\right)\right)$

Forward lens correction: k_p

Defining frame buffers: `glGenFrameBuffers`

Rendering to frame buffers:

```
glBindFramebuffer(GL_DRAW_FRAMEBUFFER, 0);
glBegin(GL_QUADS)
        glTexCoord2f(0, 0); glVertex2i(0, 0);
        glTexCoord2f(1, 0); glVertex2i(1, 0);
        glTexCoord2f(1, 1); glVertex2i(1, 1);
        glTexCoord2f(0, 1); glVertex2i(0, 1);
glEnd(GL_QUADS)
```

Projecting:

```
glBlitFrameBuffer(...);
glFinish(...);
```

Figure 3.2: Pattern projection and rendering pipeline. Patterns are computed, forward lens corrected, rendered to offscreen framebuffers, and projected in sequence very fast by blitting the buffers.

involved in software triggered acquisition.

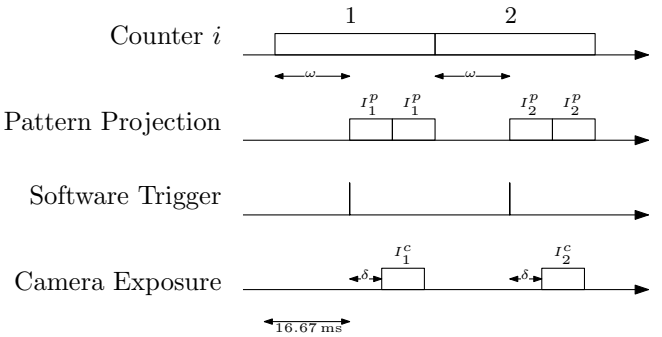


Figure 3.3: Timing diagram of software triggered pattern projection and image capture. ω denotes the temporal latency of pattern projection. δ is the camera software trigger latency. Latency lengths are only exemplary. In order to capture images correctly, each pattern is projected twice, and captured once with an exposure time of 8.33 ms.

As is clear from Fig. 3.3 the software triggered pipeline is not entirely efficient in time usage, and hence is not suitable for real time performance. The latency periods can be hidden by overlapping rendering of one pattern with camera exposure of the previous pattern. This generally requires hardware triggering the camera or projector, as the software itself does not synchronise them. Such hardware triggering is available in some pattern projectors or in the v-sync signal of display interfaces such as VGA or DVI. A timing diagram using hardware triggering and overlapped projection and exposure is shown in Fig. 3.4.

Achieving robust hardware synchronisation is more difficult than software triggering. This is because the overlapped exposure must be carefully adjusted such that projected patterns are captured faithfully, and the beginning of a particular pattern sequence is known in the captured frames. Any missed frames or missed rendering periods can easily lead to loss of synchronisation.

3.4 Precomputations for Point Triangulation

Plane-based calibration and triangulation (Sec. 2.4.2) has been used in previous real time SL systems due to its computational efficiency [WZO11]. It can determine a relative height from a pre-calibrated reference plane without the need for solving a linear system in each pixel [ZS94]. This technique is cumbersome

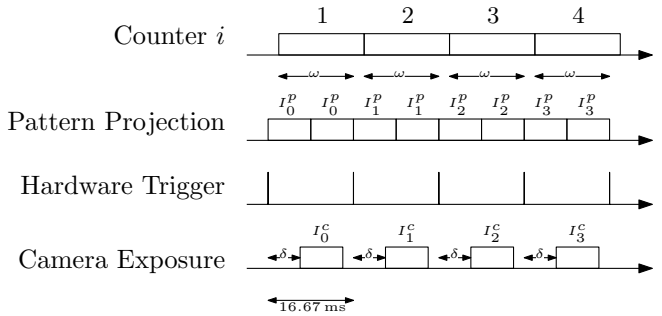


Figure 3.4: Timing diagram of hardware triggered pattern projection and image capture. ω denotes the temporal latency of pattern projection. δ is the camera hardware trigger latency. In order to capture images correctly, each pattern is projected twice, and captured once with an exposure time of 8.33 ms.

however in that a reference needs to be scanned with each calibration, and the noise from this acquisition propagates. Further, for many algorithms, the point measurements are best expressed in the camera's coordinate system.

A better alternative would be to perform direct algebraic triangulation according to Eqn. 2.14. Naively solving is prohibitive however, if real time performance is seeked in a structured light system. Restating Eqn. 2.14 for point triangulation,

$$\begin{bmatrix} \boldsymbol{P}_c(1) - u_c \boldsymbol{P}_c(3) \\ \boldsymbol{P}_c(2) - v_c \boldsymbol{P}_c(3) \\ \boldsymbol{P}_p(1) - u_p \boldsymbol{P}_p(3) \end{bmatrix} Q = 0 \quad , \tag{3.1}$$

it is apparent that after calibration, only u_p in the last row of the left-hand-side changes between scans. This allows one to pre-compute large parts of the solution, as shown by Valkenburg and McIvor [VaM98]. Defining the following determinant

$$\boldsymbol{C}_{i,j,l}^k = \det \left(\begin{bmatrix} \boldsymbol{P}_c(i) \\ \boldsymbol{P}_c(j) \\ \boldsymbol{P}_p(l) \\ \boldsymbol{I}(k) \end{bmatrix} \right) \quad , \tag{3.2}$$

where $\boldsymbol{I}(k)$ is the $k-th$ row of a 4×4 identity matrix, the coordinates of $Q = [Q_1, Q_2, Q_3, Q_4]^\top$ may be computed as

$$Q_k = C_{1,2,1}^k - u_c C_{3,2,1}^k - v_c C_{1,3,1}^k - u_p C_{1,2,2}^k + u_c u_p C_{3,2,2}^k + u_p v_c C_{1,3,2}^k \quad . \tag{3.3}$$

As seen from Eqn. 3.3, after pre-computation of C, Q can be found by means of three multiplications and three additions. This solves the computational issue with point triangulation and allows for much faster triangulation than naively solving the linear equation at each pixel.

3.5 System Design

As part of this thesis work, a complete real time SL system was constructed with the hardware trigger capabilities described above. A picture of this setup is shown in Fig. 3.5. Here the projector unit (Wintech LightCrafter4500Pro) is outputting a trigger signal at the beginning of each pattern exposure. The software performs fast point triangulation by means of the determinant tensor (Eqn.3.2), and is optimised for high performance by means of parallel computation. The overall software architecture is shown in Fig. 3.6. Example results acquired with this system are shown in Fig. 3.7. [Contribution C] of this thesis provides details of the software system.

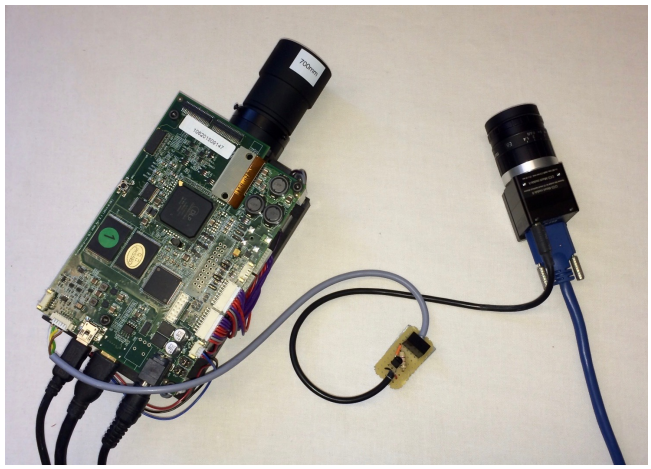


Figure 3.5: Hardware triggered structured light setup with a projector (Wintech LightCrafter4500Pro) and an industrial camera (Ximea MQ013RG). The trigger cable contains a level shifting circuit to translate the projector's 5 V trigger signal into 18 V as required by the camera.

3.6 Motion Artifacts and Correction

The fundamental assumption in scene coding with SL is that pixel-correspondence exists within the camera frames of one sequence. This is not necessarily the case, and even with small motion between frames, serious artifacts can result in the point cloud (see Fig. 3.8). Only few solutions have thus far been proposed for the motion problem. One direction of alleviating motion artifacts is to decrease

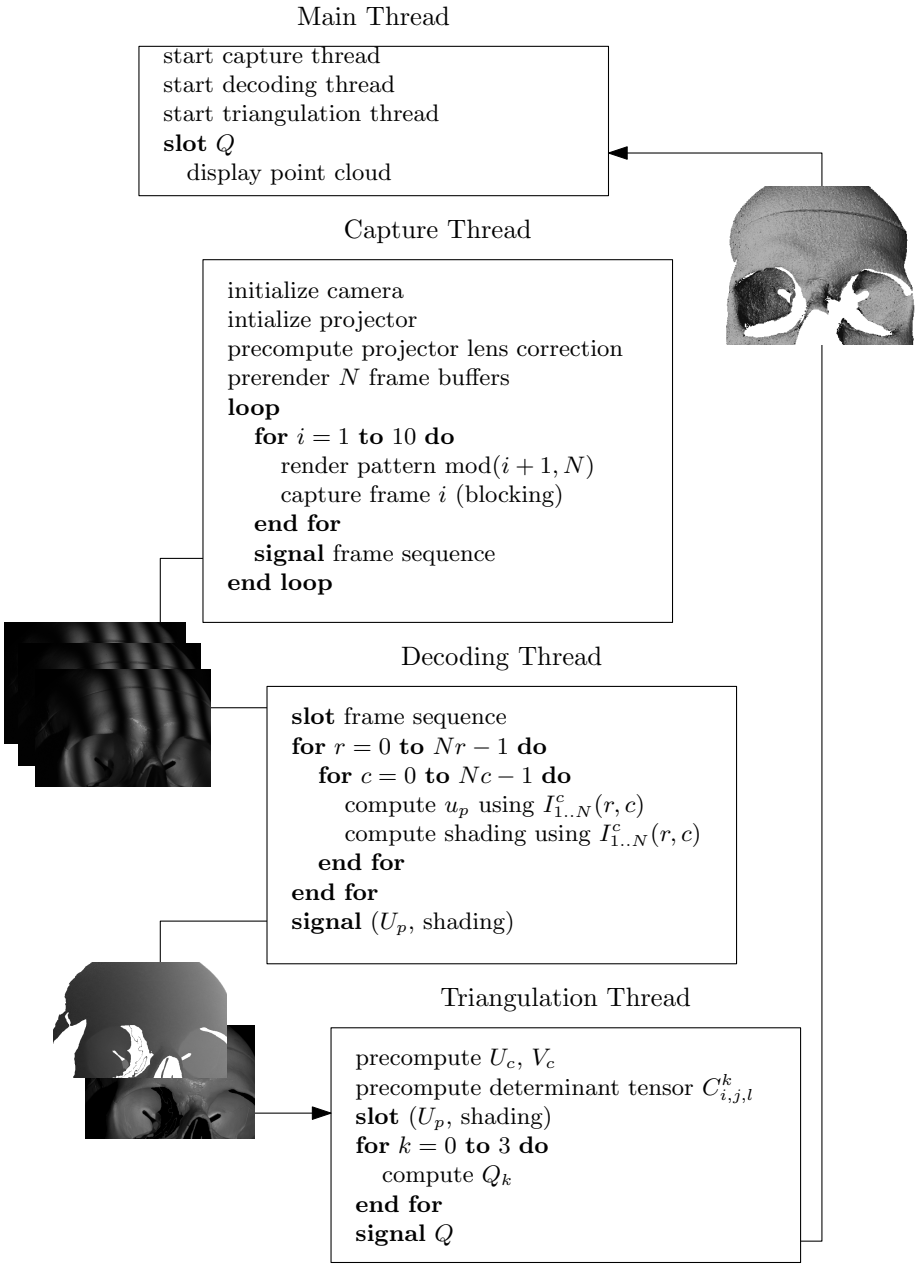


Figure 3.6: Diagram of the processing architecture of our real time SL system. Arrows indicate communication between threads by means of a signal/slot mechanism.

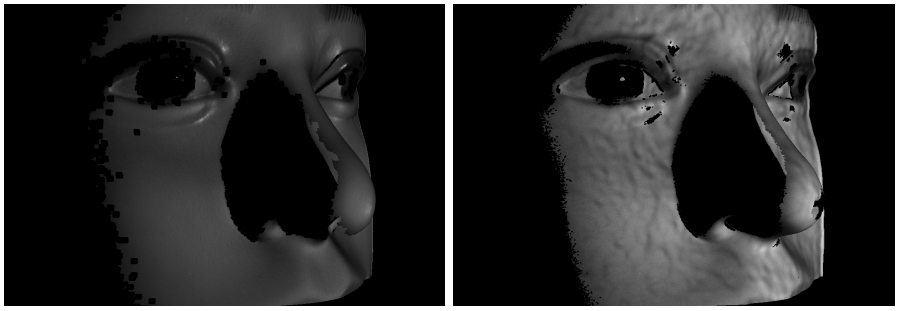


Figure 3.7: Result obtained with real time SL on a mannequin head scene. Left: 10 Hz point clouds with multi-period PS and temporal unwrapping (6 patterns). Right: 20 Hz point clouds with the 3-step PS coding method. Some spatial smoothing was applied. In both cases, quad meshes were computed for visualisation by neighbour-connection in the organised grid (see Sec. 4.1.2), hence vertices are original points.

the number of patterns or increase the pattern projection frequency, as this reduces the encoding time during which motion can occur. A high speed system of 1000 Hz which is capable of imaging a beating rabbit heart was demonstrated by Wang et al. [WLEZ13]. However, this puts additional demands on the hardware, which are difficult to satisfy. With higher frame rate and shorter exposure times, light intensity must increased accordingly, and a high-speed cameras employed. Additionally, in certain applications such as the one presented in Ch. 5, eye safe illumination intensity is a limiting factor.

Rusinkiewicz et al. use binary coding, and track pattern boundaries in order to compensate for motion during the sequence [RHHL02]. Liu et al. propose a motion corrected binary pattern method, but also employ a high speed camera and pattern projector [LGG+14]. Weise et al. provide an analysis of the error encountered in PS based SL, and derived an image based correction strategy [WLV07].

Lu et al. showed the feasibility of motion correction by means of image alignment [LXYG13]. In [Contribution E], this idea of image registration is employed by using very efficient phase correlation on a modified PS based SL sequence. This reduces motion related artifacts considerably, as seen in Fig. 3.8.

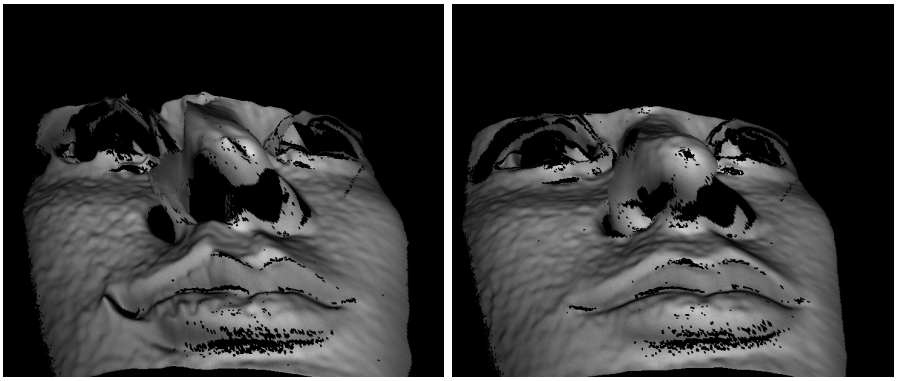


Figure 3.8: Motion artifacts in SL and their correction. Left: SL scan of a moving mannequin head with the 3-phase coding method. Right: after applying the phase-based motion correction as presented in [Contribution E]. Visualising was done as in Fig. 3.7.

3.7 Contributions

This chapter pointed to the opportunities and challenges and limitations of real time SL. The thesis contributes to the development of this technique with the open source real time SL software "SLStudio", which provides a large number of fast coding schemes, OpenGL based rendering of patterns and fast point triangulation. The software is presented in [Contribution C], and available at `http://github.com/jakobwilm/slstudio`.

A method for the reduction of motion artifacts in dynamic scenes is a largely unsolved problem. A novel method based on image registration is the subject of [Contribution E]. It uses efficient phase correlation based registration on camera frames with a modified PS method.

Chapter 4

Rigid Tracking using Structured Light

Real time SL enables the recording of dynamic phenomena in 3-D. An important class of problems is the tracking of rigid objects over time. This may be used e.g. for robotic object pose estimation, mapping, or tracking. Such tracking information is used e.g. in the clinical application described in Ch. 5.

When control feedback is involved, the tracking result must be available shortly after data acquisition, which puts strict latency demands on the tracking algorithm as well. In the following, it is shown how such real time tracking can be performed, filtered to yield higher accuracy and evaluated using an alternative tracking device.

4.1 Rigid Alignment

The foundation of tracking with SL is the alignment, or registration, of point clouds. Considering a reference point cloud $\{R\}$ to which data point cloud $\{D\}$, the alignment process estimates a rigid transformation which brings $\{D\}$ into alignment with $\{R\}$. Using a rotation matrix $\boldsymbol{R}_\mathrm{a}$ and translation vector $\boldsymbol{T}_\mathrm{a}$ to parameterise this transformation, $\boldsymbol{R}_\mathrm{a}\{D\} + \boldsymbol{T}_\mathrm{a}$ should be similar to $\{R\}$, possibly with only partial overlap.

4.1.1 Literature Overview

The alignment of point clouds can be based on sparse point features or directly on all available data. With point features, a small number of points are detected in both point clouds and matched robustly [AL10]. Pose tracking using regression forests was proposed in [FWGV11]. In [BKW+08], head tracking is based on nose-detection. As these model based methods were trained on a large amounts of sample data, they are robust, but not particularly precise. Other alternatives consist in the registration of a sparse set of 3D geometry features, such as spin images [Joh97] or signed distance fields [Mas01]. A general review of point cloud registration methods is given in [SMFF07].

Using the full point cloud for registration makes use of all available data, and as such is often superior when maximum accuracy is needed [SMFF07]. Therefore, many applications requiring accuracy, e.g. robot localisation, are based on Iterative Closest Point (ICP) and similar algorithms [LM94, SHK94, NDI+11]. Some modern alternatives exist for dense rigid alignment, including the "Normal Distributions Transform" [Mag09] which represent the point cloud as a piecewise smooth function. The method of Bing and Baba constructs intermediate Gaussian mixture model representations to be be registered [JV11]. While these methods were shown to be very powerful and more robust, especially in case of noisy point clouds and bad initial alignment, they are computationally heavier than the ICP.

The ICP algorithm is the most well-established method for the dense alignment of point clouds. It iterates over the process of finding nearest neighbour correspondences and alignment by finding the least square rigid transformation, also known as the orthogonal Procrustes solution. ICP has been used extensively since its introduction in the early 1990's. The original method of Chen and Medioni [CM91] minimised a linearised point-to-plane metric, while the concurrent paper of Besl and McKay [BM92] minimises point-to-point distances and proposes an extrapolation strategy for faster convergence. A wealth of literature has proposed improvements to the original ICP. These include the use of efficient normal space based sampling [RL01], rejection of boundary matches [TL94] (see Fig. 4.1) or the use of robust estimators such as the least median square [TFR99] or least trimmed squares [CSK05]. It is also possible to forego the iterative procedure of ICP, and find the optimal transformation directly by means of non-linear optimisation [Fit03]. A comparative analysis of many of these techniques is provided in [RL01]. Another comparison in the context of head pose tracking was performed by the thesis author in [WOP+11].

The main computational complexity in ICP lies in the determination of nearest neighbours. For this reason, spatial search structures such as octrees of kD-trees

are usually employed. In cases where point data is organised in an ordered grid, the matching can be done very efficiently using "reverse calibration" [BL95] as will be shown later.

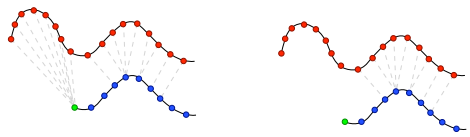


Figure 4.1: Concept of edge rejection in the ICP algorithm. Point pair including the query edge are rejected as proposed by [TL94] which allows registration of point clouds with partial overlap.

4.1.2 Fast ICP for Tracking

Tracking by means of ICP involves defining a reference point cloud to which others are aligned. Choosing the first point cloud (index $i = 0$) allows to compute the necessary search structure, normals (necessary for point-to-plane), and boundaries for "edge rejection" (as seen in Fig. 4.1) once, and to reuse these data structures when the current point cloud is to be registered to the reference. A good initial alignment for the registration at $i = n$ is the result of the registration at $i = n - 1$.

In order to align point clouds of approximately 300 000 points in real time, according to the definition of real time SL on page 23, even look-ups in spatial search structures can be prohibitive [WOP+11]. As mentioned, the task of correspondence finding is computationally demanding. It is possible to exploit the fact that point clouds from SL scanners have an inherent spatial organisation according to the regular camera pixel grid in which they are reconstructed. Such point clouds have been termed "organised point clouds" [HRD+12], and are different from "depth maps" in that individual (x, y, z) coordinates exist at each pixel position (u_c, v_c).

In order to identify edge points in organised point clouds, a simple traversal over the points is sufficient. Any point whose 8-neighbourhood in the grid contains empty values can be marked as an edge point. Normal estimation can also be performed very efficiently using the "integral image" approach of Holzer [HRD+12]. Finally, point matching can be done using the "reverse calibration" technique. Considering reference point cloud $\{R_{u,v}\}$ to which data point cloud $\{D_{u,v}\}$ is to be registered, both of which are expressed in the camera with matrix $\boldsymbol{P}_c$. The point match for point $D_{u0,v0}$ is established by computing

its projection

$$\boldsymbol{d}_{u0,v0} = [p_x, p_y, s]^\top = \boldsymbol{P}_c D_{u0,v0} \quad , \tag{4.1}$$

and matching it to

$$R_{\text{round}(p_x/s),\text{round}(p_y/s)} \quad , \tag{4.2}$$

if this point exists, and is not an edge point. While the quality of matches obtained in this way might not be as high as with Euclidean nearest neighbours, the speed with which they are determined is much higher. Since the estimated transformation in tracking is usually small, this algorithm generally converges after few iterations. Fig. 4.2 shows a tracking scene of a mannequin head which was mounted on a motor performing a periodic movement.

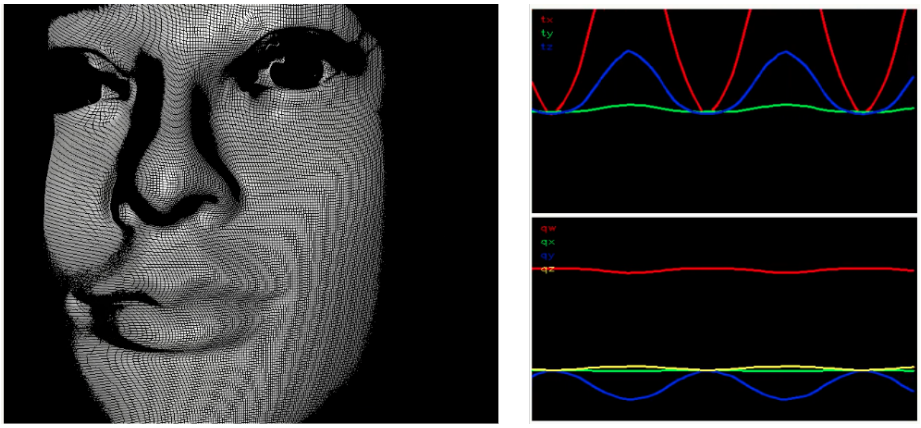


Figure 4.2: Tracking based on SL and the fast ICP method described in Sec. 4.1.2. A mannequin head is mounted to a stepper motor and rotated back and forth. Result are computed in real time and displayed on screen.

4.2 Parameterisation

A rigid pose has six degrees of freedom, and may be described using a number of parameterisations, e.g. translations + Euler angles, translations + rotation matrix, dual quaternion, or translations + unit quaternion. Euler angles parameterise the rotation using the minimum number of three parameters , but have singularities, and their math does not lend to easy formulation of filtration, composition or extrapolation. Rotation matrices are readily composed, but not extrapolated, and they are cumbersome in many analytical derivations,

as their special properties must be retained (orthogonality and positive unit determinant). Most of these issues are alleviated with quaternions. While the dual-quaternion representation combines translation and rotation, and thereby provides a unified algebra for roto-translations, most of these benefits can be obtained with a combined translation vector + unit quaternion, which has better interpretability. The representation of the current pose used here is thus as follows:

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{q} \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \\ t_z \\ q_w \\ q_x \\ q_y \\ q_z \end{bmatrix} \quad , \|\boldsymbol{q}\| = 1 \quad , \tag{4.3}$$

where $\boldsymbol{t}$ is the translation and $\boldsymbol{q}$ the quaternion. While the quaternion has four degrees of freedom, its length is restricted to unity, thereby avoiding over-parameterisation. The quaternion has its own specific algebra [Kui02], but can easily be converted to other representations of rotation. For instance, the equivalent axis angle representation is $\left(\left[\frac{q_x}{\sqrt{1-q_w q_w}}, \frac{q_y}{\sqrt{1-q_w q_w}}, \frac{q_z}{\sqrt{1-q_w q_w}} \right]^\top, 2\arccos(q_w) \right)$ while a rotation matrix can be computed as

$$\boldsymbol{R} = \begin{bmatrix} 1-2q_y^2-2q_z^2 & 2q_xq_y-2q_zq_w & 2q_xq_z+2q_yq_w \\ 2q_xq_y+2q_zq_w & 1-2q_x^2-2q_z^2 & 2q_yq_z-2q_xq_w \\ 2q_xq_z-2q_yq_w & 2q_yq_z+2q_xq_w & 1-2q_x^2-2q_y^2 \end{bmatrix} \quad . \tag{4.4}$$

Hence, the quaternion representation can be used throughout, but converted to other representation where such are more convenient.

4.3 Kalman Filtering

The tracking method described so far is fast and robust. However, the tracking result can contain noise. A natural question is whether physical constraints or priors can be applied such that noise is reduced and tracking precision increased. The Kalman filter is a hallmark of signal processing since the 1960's [Kal60], and provides a framework for filtering signals and sensor fusion with priors on the signals covariance structure. This allows e.g. in GPS based tracking to include constraints on a vehicle's acceleration and speed, and improve the precision of the momentary positional estimate.

The Kalman filter is based on the assumption of white noise on measurements. If this assumption is not met, the filter may diverge or be very limited in perfor-

mance . Additionally, due to its linear formulation, the model needs to be linear as well. If that is not the case, linearisation can be performed, as will shown below. In that case, the filter is called the Extended Kalman Filter (EKF) [EW99]. Alternatives exist, if linearisation is not possible, or unreliable, such as particle filters [AMGC02], but these are generally much more computationally demanding.

The Kalman filter formulates a measurement vector, $\boldsymbol{z}_k$, with the observable measurements:

$$\boldsymbol{z}_k = h(\boldsymbol{s}_k) + \boldsymbol{\psi}_k \quad , \tag{4.5}$$

where $h(\cdot)$ is the so-called measurement function, and $\boldsymbol{\psi}_k$ a "measurement-noise" vector, considered uncorrelated in time with covariance $\boldsymbol{\Psi}$. k denotes an index, often a time-point. $\boldsymbol{s}_k$ is a state vector, which may contain observable properties as well as non-observable properties. It is formulated as a propagation from the last state by means of a state-transition function $\phi(\cdot)$, and possibly a known alteration, or control vector $\boldsymbol{u}_k$, which affects the state through $b(\cdot)$:

$$\boldsymbol{s}_k = \phi(\boldsymbol{s}_{k-1}) + b(\boldsymbol{u}_k) + \boldsymbol{\theta}_k \quad . \tag{4.6}$$

$\boldsymbol{\theta}_k$ is an addition "process noise", considered independent white noise with covariance matrix $\boldsymbol{\Theta}$.

Given measurements which can be formulated in this way, the Kalman filter can provide statistically optimal estimates of the "true" state. It should be noted that the requirements of the filtering framework are:

- That the phenomenon is fully modelled, e.g. that $h(\cdot)$ and $\phi(\cdot)$ faithfully describe the physics underlying the state and measurement.
- That $\boldsymbol{\psi}_k$ and $\boldsymbol{\theta}_k$ are truly uncorrelated in time (white noise). This does not entail any requirements on the amplitude distribution other than an expected value of 0 and finite variance.
- That the covariance-structures $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ are given or can be estimated.

The Kalman filter performs two steps:

(1) Prediction, in which the current state and its estimated covariance is propagated, and optionally a control input is applied to yield a *a-priori* state estimate.

(2) Correction, where the current measurement is involved, to yield a statistically optimal *a-posteriori* estimate.

The seminal paper of Kalman [Kal60] describes a linear filter, in which $h(\cdot)$, $\phi(\cdot)$, and $b(\cdot)$ are linear functions, i.e. can be applied through matrix-vector multiplication with matrices $\boldsymbol{H}$, $\boldsymbol{\Phi}$ and $\boldsymbol{B}$ respectively. In that case, the prediction and correction steps have the following closed-form solutions:

(1) prediction:

$$\boldsymbol{s}_k^- = \boldsymbol{\Phi}\boldsymbol{s}_k + \boldsymbol{B}\boldsymbol{u}_k \tag{4.7}$$

$$\boldsymbol{P}_k^- = \boldsymbol{\Phi}\boldsymbol{P}_{k-1}\boldsymbol{\Phi}^\top + \boldsymbol{\Theta} \tag{4.8}$$

(2) correction:

$$\boldsymbol{K}_k = \boldsymbol{P}_k^- \boldsymbol{H}^\top \left(\boldsymbol{H}\boldsymbol{P}_k^- \boldsymbol{H}^\top + \boldsymbol{\Psi}\right)^{-1} \tag{4.9}$$

$$\boldsymbol{s}_k = \boldsymbol{s}_k^- + \boldsymbol{K}_k \left(\boldsymbol{z}_k - h(\boldsymbol{s}_k^-)\right) \tag{4.10}$$

$$\boldsymbol{P}_k = \left(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}\right)\boldsymbol{P}_k^- \tag{4.11}$$

Matrix $\boldsymbol{P}_k$ is the covariance of the estimate, and the matrix $\boldsymbol{K}_k$ the so-called Kalman-gain, which defines the the extend to which the current measurement influences the state estimate.

In the more general case of non-linear functions $h(\cdot)$, $\phi(\cdot)$, and $b(\cdot)$, these functions can be linearised locally, and substituted by their Jacobians in the relations above.

In the pose signal filtering problem, the measurement vector is equal to the pose parameterisation, 4.3. There is no control-input, and hence $b(\boldsymbol{u}_k) = 0$. In order to model the distribution of linear and angular velocities, the following state vector is formed:

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{q} \\ \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix}_{13\times 1} , \tag{4.12}$$

where $\boldsymbol{t}$ is the translation vector, $\boldsymbol{q}$ the unit rotational quaternion, and $\boldsymbol{v}$ and $\boldsymbol{\omega}$ are 3×1-vectors of the current linear and angular velocities.

4.3.1 Linearised Quaternion Propagation

The functions involved in the state transition and measurement of the pose signal parameterised using Eqn. 4.3 shall now be linearised, with the internal state as given in Eqn. 4.12. First, it is useful to relate the quaternion to angular velocity vector $\boldsymbol{\omega}$. Naturally, these can be obtained as the time derivates as

follows [GA98]:

$$\frac{\partial \boldsymbol{q}}{\partial t} = \frac{1}{2}(0, \omega_x, \omega_y, \omega_z)^\top \boldsymbol{q} = \frac{1}{2}\tilde{\boldsymbol{\Omega}}\boldsymbol{q} \quad . \tag{4.13}$$

Quaternion propagation can then be done in the following way:

$$\boldsymbol{q}_k = \exp\left(\Delta t \tilde{\boldsymbol{\Omega}}\right) \boldsymbol{q}_{k-1} = \left[\cos\left(\frac{\|\omega\|\,\Delta t}{2}\right)\boldsymbol{I} + \frac{2}{\|\omega\|}\sin\left(\frac{\|\omega\|\,\Delta t}{2}\right)\tilde{\boldsymbol{\Omega}}\right] \boldsymbol{q}_{k-1} \quad , \tag{4.14}$$

and the full state transition function function becomes

$$\phi(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{t} + \Delta t \boldsymbol{v} \\ \exp\left((\Delta t \tilde{\boldsymbol{\Omega}})\boldsymbol{q}\right) \\ \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix} \quad . \tag{4.15}$$

Since the measurements are simply part of the state, the measurement function is very simple. It also conveniently takes care of quaternion re-normalisation:

$$h(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{t} \\ \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|} \end{bmatrix} \tag{4.16}$$

The Jacobian of $\phi(\cdot)$ is defined as follows:

$$\boldsymbol{\Phi} = \begin{bmatrix} \frac{\partial \phi_{t_x}}{\partial t_x} & \cdots & \frac{\partial \phi_{t_x}}{\partial \omega_z} \\ \vdots & & \vdots \\ \frac{\partial \phi_{\omega_z}}{\partial t_x} & \cdots & \frac{\partial \phi_{\omega_z}}{\partial \omega_z} \end{bmatrix}_{13\times 13} . \tag{4.17}$$

Deriving these components yields [GA98]:

$$\frac{\partial \phi_{t_i}}{\partial v_i} = \Delta t \tag{4.18}$$

$$\frac{\partial \phi_{\boldsymbol{q}}}{\partial \boldsymbol{q}} = \exp\left(\Delta t \tilde{\boldsymbol{\Omega}}\right) \tag{4.19}$$

$$\begin{aligned} \frac{\partial \phi_{\boldsymbol{q}}}{\partial \omega_i} &= \frac{\partial \exp\left(\Delta t \tilde{\boldsymbol{\Omega}}\right) \boldsymbol{q}}{\partial \omega_i} \\ &= \left[-\frac{\omega_i \Delta t}{2\|\boldsymbol{\omega}\|}\sin\left(\frac{\|\boldsymbol{\omega}\|\Delta t}{2}\right)\boldsymbol{I}_{4\times 4}\right. \\ &\quad + \left(\frac{\omega_i \Delta t}{\|\boldsymbol{\omega}\|^2}\cos\left(\frac{\|\boldsymbol{\omega}\|\Delta t}{2}\right) - \frac{2\omega_i}{\|\boldsymbol{\omega}\|^3}\sin\left(\frac{\|\boldsymbol{\omega}\|\Delta t}{2}\right)\right)\tilde{\boldsymbol{\Omega}} \\ &\quad \left. + \frac{1}{\|\boldsymbol{\omega}\|}\sin\left(\frac{\|\boldsymbol{\omega}\|\Delta t}{2}\right)\frac{\partial \tilde{\boldsymbol{\Omega}}}{\partial \omega_i}\right]\boldsymbol{q} \end{aligned} \tag{4.20}$$

Similarly for the measurement function $h(\cdot)$, we have

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial h_{t_x}}{\partial t_x} & \cdots & \frac{\partial h_{t_x}}{\partial q_z} \\ \vdots & & \vdots \\ \frac{\partial h_{q_z}}{\partial t_x} & \cdots & \frac{\partial \phi_{q_z}}{\partial q_z} \end{bmatrix}_{7\times 13} , \tag{4.21}$$

whose relevant components are more easily derived as

$$\frac{\partial h_{\boldsymbol{t}}}{\partial \boldsymbol{t}} = \boldsymbol{I} \qquad \frac{\partial h_{q_i}}{\partial q_i} = \frac{\|\boldsymbol{q}\|^2 - q_i}{\|\boldsymbol{q}\|^3} \tag{4.22}$$

The above filtering approach was used on data of the motorised mannequin head setup. An example of the filtered pose signal and its unfiltered counterpart is shown in Fig. 4.3. The covariances $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ were set using reasonable values.

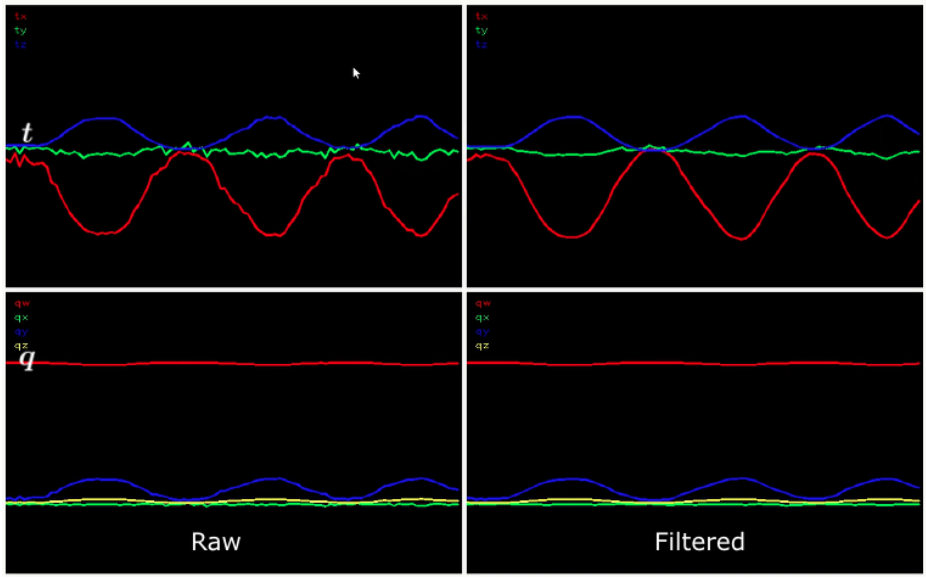


Figure 4.3: Results of applying the EKF on a pose signal obtained online on a rotating mannequin head.

A noticeable reduction of noise is seen after applying the EKF and satisfactory qualitative results. In the next section, the tracking and filtering approach will be validated by means of a secondary tracking device.

4.4 Coordinate Alignment

Since it is difficult to establish ground truth in a tracking experiment, in order to evaluate tracking performance, a secondary, accurate device can be used. It is not trivial however to perform such evaluation, as the tracking result will be expressed in different coordinate systems. Fig. 4.4 shows a tracking experiment on a mannequin head with SL based tracking as described above, and a commercial marker-based tracker (Polaris Vicra).

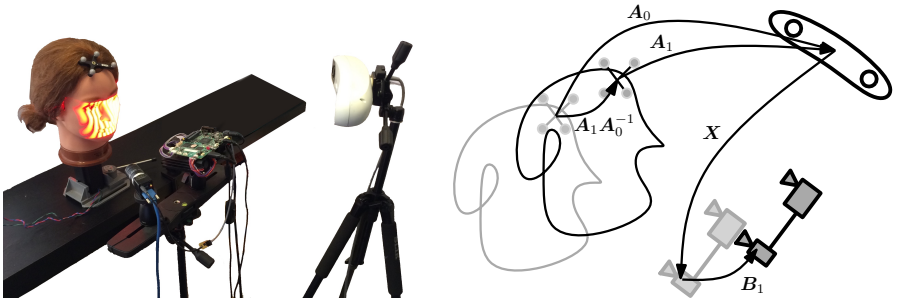


Figure 4.4: Head tracking experiment. Left: the real time SL system scanning, while the head pose is also measured using the Polaris Vicra tracker with the tracking tool attached to the top of the mannequin head. Right: diagram of the coordinate transformation between both tracking systems, as used in hand-eye calibration.

With both systems observing transformations of the mannequin head in different coordinate systems, the relation between coordinate systems may be found by means of "hand-eye calibration". Referring to Fig. 4.4 (Right), the current optical marker pose is reported as $\tilde{\boldsymbol{A}}_i$, which is a 4×4 homogenous matrix containing rotation and translation. The SL tracker determines each pose relative to the pose at time $t = 0$ as $\tilde{\boldsymbol{B}}_i$. The ideal relation between these pose data is given by rigid transformation $\boldsymbol{X}_{4\times4}$, where the following is satisfied

$$\left(\tilde{\boldsymbol{A}}_i \tilde{\boldsymbol{A}}_0^{-1}\right) \boldsymbol{X} = \boldsymbol{X} \left(\tilde{\boldsymbol{B}}_i \tilde{\boldsymbol{B}}_0^{-1}\right) \tag{4.23}$$

$$\boldsymbol{A}_i \boldsymbol{X} = \boldsymbol{X} \boldsymbol{B}_i \quad . \tag{4.24}$$

Determining $\boldsymbol{X}$ based on a number of such relative motions is referred to as hand-eye-calibration. A number of closed form solutions exist for this problem. Many of these methods solve for the optimal rotation first, then for the translation, e.g. [TL89], others are simultaneous, and often based on parameterisations with quaternions [Dan99]. A qualitative comparison of many of these solutions is presented in [SEH12].

The hand-eye calibration has the fundamental shortcoming that rotations and translations have different inherent scales, and are not directly comparable. Hence, the algebraic error minimised in these methods does not necessarily bring 3-D points into the best possible correspondence. Hence, for the tracking experiment shown in Fig. 4.4, the hand-eye calibration according to [TL89] was used as a closed form initialiser to a non-linear optimisation procedure. In this procedure, a surface point cloud of the mannequin head was used, and the Root Mean Square Error (RMSE) of tracking as the error objective. Results after this coordinate alignment are shown in Fig. 4.5. The obtained RMSE on tracking deviations was below 2 mm in all cases. These results show successful rigid object tracking using our approach.

4.5 Contributions

As was shown in this chapter, rigid pose tracking can be done fast and accurately by means of SL, projective ICP and the EKF. The thesis contributes to SL based rigid tracking by implementing of real time efficient rigid alignment for streaming SL data as detailed above. [Contribution A] discusses part of the methods described above, albeit based on a TOF sensor instead of SL. A scheme for coordinate system alignment for the evaluation of tracking results using hand-eye-calibration and optimisation was also described in this chapter, and it verified the correctness and accuracy of tracking results.

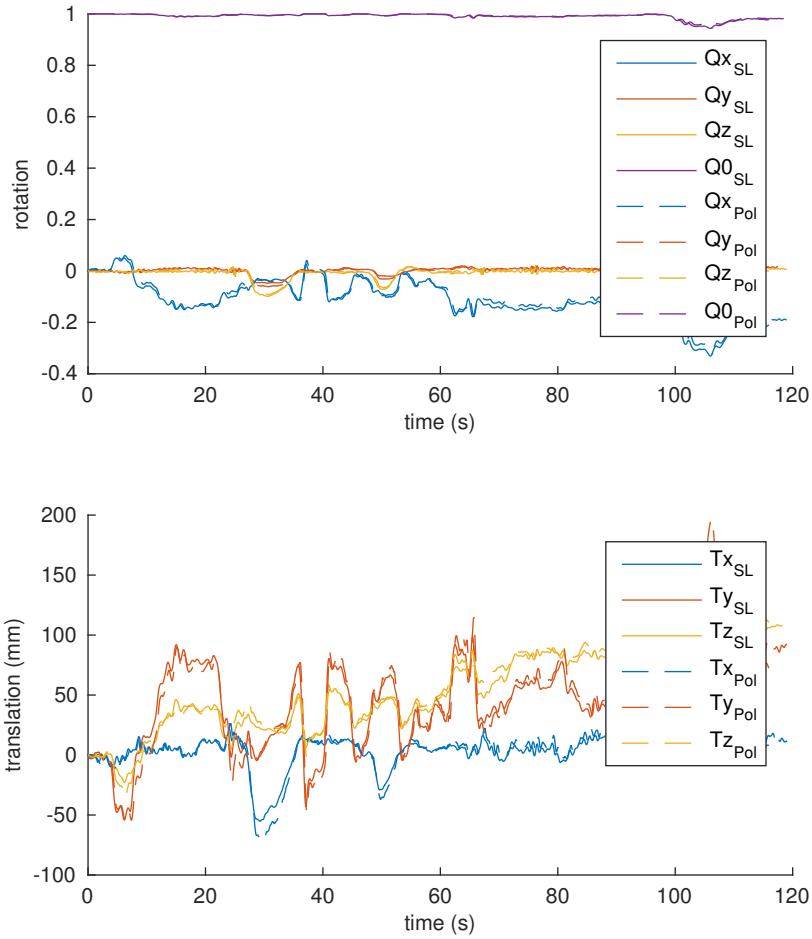


Figure 4.5: Result of tracking experiment on a mannequin head that was moved freehand and tracked using both the Polaris Vicra marker based system and the SL based tracking strategy. Coordinate systems were aligned using hand-eye calibration followed by RMSE-based optimisation.

Chapter 5

Medical Motion Correction using Structured Light

Medical brain-scanners, including MRI, Positron Emission Tomography (PET) and Computed Tomography (CT) have ever increasing resolution and fidelity. This enables doctors to perform better diagnosis on diseased patients. In addition, it makes possible a number of new techniques to improve the study of brain activity and the neurological system. The benefits of increased resolution are unfortunately dissolved in part due to patient motion while the scan is on-going. When gross motion occurs, regular artifacts appear in the images, as seen in Fig. 5.1. The same issues arise in radiation therapy, where a high-precision and high-intensity photon beam is directed at cancerous tissue to destroy it. Patient motion during this procedure leads to a higher, unneeded radiation dose. The traditional methods of avoiding these issues consists of head immobilisation with various devices, including thermoplastic face masks, vacuum pillow (see Fig. 5.2), or even anaesthesia in un-cooperative patients or children.

Motion correction for medical imaging is a broad field with many approaches suggested in literature. These methods can be categorised as either retrospective (data is corrected offline after acquisition) and prospective (scan parameters are adjusted or data is corrected on the fly). A different separation is in image based methods, and methods based on external tracking devices, such as the one that will be described in this chapter. Depending on the primary modality, both retrospective and prospective motion correction methods have their merits.

In PET, a detector ring records gamma rays from positron-emitting tracers. Both static recordings of metabolic activity and dynamic studies of tracer uptake are conducted. Motion correction can be based on time framed reconstructions, known as Multiple Acquisition Frames (MAF) [PT97]. This was done using an external optical tracking device by Fulton et al. [FME+02]. This however limits

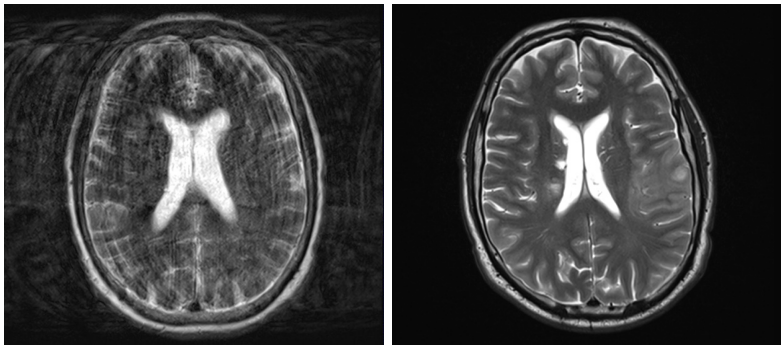


Figure 5.1: Example of MRI brain scan (Siemens mMR T_2 contrast). Left: with patient motion the image quality is severely reduced. Right: without patient motion the diagnostic value of the image is much higher.

Figure 5.2: Methods of head immobilisation for radiotherapy and motion-free medical scans. Left: vacuum pillow immobilising an infant. (c) Dept. of Pediatric Radiology, Wilhelmina Children's Hospital, UMC Utrecht - Utrecht/NL. Right: thermoplastic mask for head immobilisation. (c) MRI-P Solution, Orfit Industries n.v..

the temporal resolution of dynamic studies, and cannot correct for fast motion. With high-speed external trackers, potentially every single signal detection can be motion corrected. This was demonstrated e.g. by Carson[CBLJ03]. Single event based correction is complex though in that radio-density and scanner geometry must be considered in the process. There is little or no benefit to prospective correction in PET, and reconstruction is always performed offline after data recording. The PhD thesis of Oline V. Olesen [Ole11] provides a review of motion correction strategies for PET and proposes SL for retrospective correction.

For CT, most often fast structural imaging is done with scan times between fractions of a second and a few minutes. Multiple recordings may be acquired with and without contrast agent, or along the body for a complete recording. Image based correction methods are demonstrated in e.g. [SZRvH05, LSYX06]. Since spatial resolution is very important, also CT can benefit from motion correction, specifically when respiratory motion is involved. Here, the "gating" technique is widely used, where data is collected only during the same phase of the cardiac or respiratory cycle [DK04, BK04].

In MRI, acquisition times can be very long, and motion is often the limiting factor for scan duration. As the MRI modality is very versatile, a wealth of methods have been proposed with various trade-offs between contrast type, image quality, scan speed, and other parameters. It has become very common to apply so-called "navigator" sequences within the MRI protocol to detect patient motion, and re-acquire data if necessary. These navigators add to the collected scan time however, which makes their use prohibitive in some MRI techniques [HMW$^+$12].

With functional imaging techniques such as functional MRI (fMRI), the very subtle "BOLD-contrast" is measured, and allows to deduce which parts of the brain are active during a certain stimulus. In diffusion MRI (dMRI), the anisotropic diffusion coefficient of water molecules is measured to non-invasively gain insight into the main fiber/axon directions within the brain. These modalities require large amounts of data, and often, long acquisition times. Their analysis is based on multiple recordings of the same tissue under different acquisition parameters, with the assumption of stationarity. Consequently, motion during measurement leads to artifacts in the reconstructed data, and possibly biased data. Prospective motion correction has numerous benefits in MRI over retrospective correction. With prospective tracking information, it is possible to adjusts magnetic field gradients, such that the MRI-scanner's internal coordinate system effectively follows the patient's head, as shown in Fig. 5.3. This avoids uneven sampling of the frequency space and complex realignment procedures [DZS$^+$06].

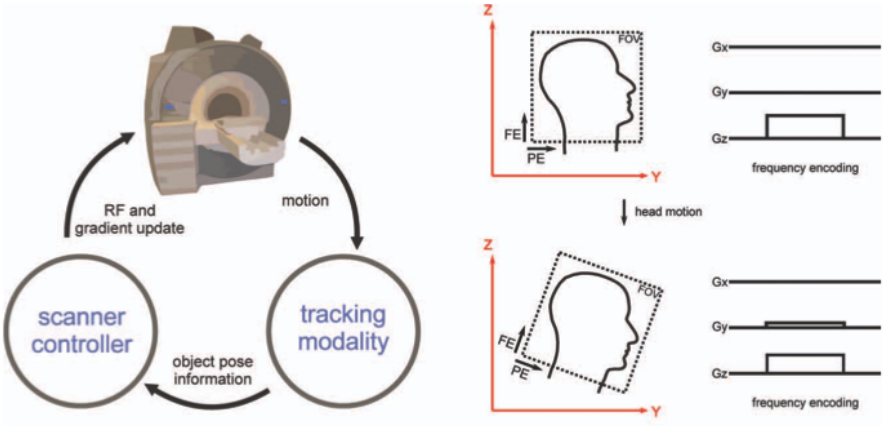


Figure 5.3: The prospective motion correction cycle in MRI. The scanner's imaging protocol is updated such that its coordinate system follows the patients head, avoiding post-process motion correction and violations of the sampling theorem. From MacLaren et al. [MHSZ13].

Forman et al. use a coded checkerboard marker and a single camera for tracking in MRI [FAHB11]. Schulz et al. presented in-bore tracking of optical markers with a single MRI-compatible camera [SSR+12]. Gholipour et al. use a magnetic field sensor attached to the patient to track motion [GPvdK+11]. The KinetiCor MRI tracker is a commercial offering which uses a single camera and marker exploiting the Moiré effect. The MoTrak system of Psychology Software Tools, Inc. uses an Radio Frequency (RF) transmitter-receiver design for MRI head motion tracking. A review of current prospective motion correction techniques for MRI is found in [MHSZ13].

Most all methods for non image-based motion tracking in medical scanners use optical or magnetic field markers. This is problematic in clinical workflows, adds to patient discomfort, and is difficult to rigidly attach to the patients skull. Notable exceptions are the SL based system of Olesen [OJP+10] and a Simultaneous Localisation and Mapping (SLAM) based passive stereo approach suggested by [KSM+14].

5.1 The Tracoline MRI System

In order to track motion for prospective motion correction in MRI without using markers, an MRI compatible version of the real time SL system discussed in Ch. 3 was developed. This is based on a similar system which was successfully used for retrospective PET motion correction [Ole11], called the "Tracoline" system.

The approach used in "Tracoline" is to acquire an initial reference point cloud of the patient's upper face, including nose and the area around the eyes. The reference point cloud may be manually edited to remove eye regions and other facial regions which exhibit non-rigid movement, such that the motion of these regions does not influence the tracking result.

5.1.1 SL through Optical Fiber Bundles

A big issue with any electronic device that is used in or close to an MRI-scanner is MRI-compatibility. MRI-compatibility can be difficult to achieve in practice for the following reasons:

- The presence of a very strong magnetic field. The field strength of clinical scanners is in the range $1.5 - 9\,\mathrm{T}$, measured in the centre of the scanner gantry, that is up to $3.6 \cdot 10^5$ times stronger than the earth's magnetic field. Due to safety precautions, this makes it impossible to have ferromagnetic objects in the vicinity of the scanner.
- Due to the scanner's reliance on magnetic field homogeneity, many electrically conductive objects cannot be used inside the gantry because inductive effects create local magnets which distort the recorded image.
- MRI-machines emit strong radio-magnetic signals, which interfere with many electronic components, e.g. camera sensors and DMDs.
- Because the MRI signal is rather weak RF-waves, any device must not produce interfering RF signals, or be shielded.

For the above-mentioned reasons, MRI examination rooms are constructed as Faraday cages.

In order to overcome the aforementioned issues, the "Tracoline" system encloses the camera and projector in a Faraday cage, which is situated in the examination

room as shown in Fig. 5.4. Coherent optical fiber bundles lead from this cage into the gantry, where camera lenses are mounted on the fiber ends. These parts are shown in Fig. 5.5. Hence, all sensitive electronic components are shielded from electromagnetic influences, and the MRI scanner is protected from their radiation. The two-view geometry resides in the scanner's gantry, and is little affected by the relay through fiber bundles (see Fig. 5.6).

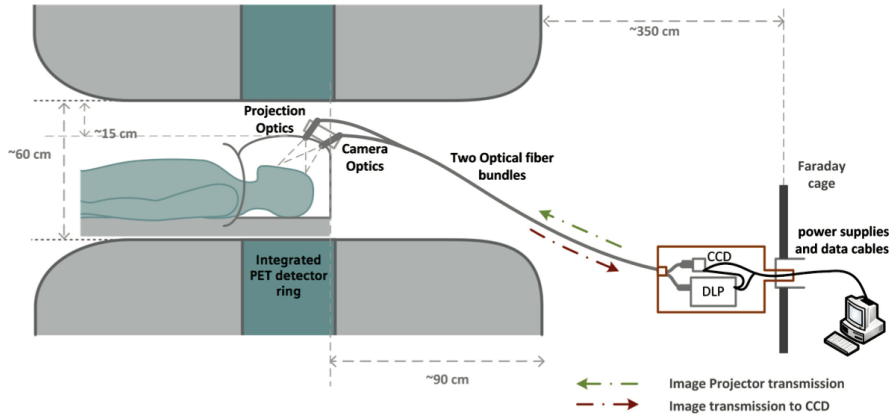


Figure 5.4: Schematic diagram of the "Tracoline" system for MRI motion correction. SL system components are located outside the scanner gantry but in the examination room in a Faraday cage. Coherent image fiber bundle used to relay images from the projector to the patients face and to the camera. Data is transmitted through optical extenders, and DC power provided through the scan room filtering panel. Courtesy of Oline V. Olesen.

5.1.2 Near-Infrared Illumination

For the "Tracoline" system, illumination at visible wavelengths would disturb patients considerably. Hence, near-infrared illumination is employed. This modification gives essentially "imperceptible" SL, which could see use in many other areas apart from this application [FSV04].

Since projectors are not commonly available with such illumination, a commercial projector (Wintech LightCrafter 4500) was retrofit with a near-infrared LED illumination source (810 nm). The mechanism of colour modulation in this DLP based projector is by means of temporal switching between the LEDs. The

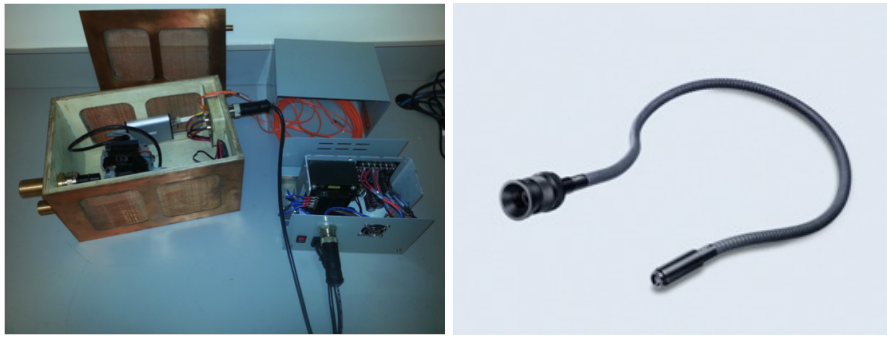


Figure 5.5: Left: Faraday cage for the "Tracoline" SL system. The box contains camera, projector and power supply. Right: Coherent image fiber bundle. (c) Schott AG.

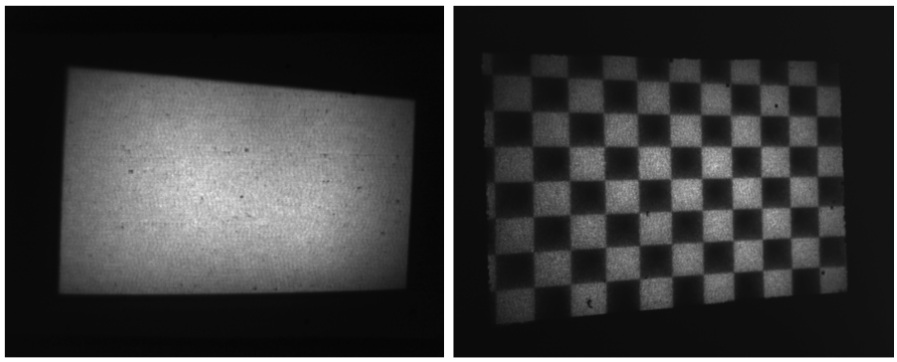


Figure 5.6: Left: image of white board captured through the coherent fiber bundle. Right: image captured of projected checkerboard – both light paths through optical fiber bundles.

light of all LEDs is combined and directed towards the micromirror array by means of dichroic mirrors, which pass certain wavelengths and reflect others. The infrared modification uses the red dichroic mirror which is reflecting light in a broad band above 600 nm. Hence the red LED can readily be swapped with one of higher wavelength and a similar voltage-current curve.

Standard silicon based camera CCDs are also only sensitive within a narrow wavelength range, covering little more than the visible range of $300 - 900$ nm. By optimising the CCD and omitting the infrared filter, silicon based CCDs can be made sensitive up to 50% quantum efficiency at 810 nm, which is sufficient for infrared SL.

5.2 Contributions

As discussed in this chapter, prospective motion correction has certain advantages over retrospective techniques in MRI. In addition, marker-less tracking is highly beneficial in clinical settings [Ole11]. With the real-time SL approach elaborated in chapters 3 and 4, rigid pose tracking can be done accurately, as also demonstrated in [Contribution A]. To solve the MRI compatibility problem, the SL system must be made compatible with MRI scanners, which is the subject of the patent in [Contribution F]. Here the idea of relaying the SL optical signal to and from the MRI gantry is detailed, and the overall design is covered.

Results of real time SL scans in MRI machines are demonstrated in [Contribution D], which demonstrates that neither the SL nor the MRI systems are affected, and that SL surface scans of human volunteers can successfully be acquired while lying inside the MRI scanner.

Chapter 6

Accuracy and Precision

An important question with any measurement system is which accuracy, precision and resolution can be achieved. With SL, there is no straight forward answer, as the attainable performance is influenced by numerous factors, including

- Calibration quality. A discussion on calibration methods is given in Sec. 2.1.
- Camera and projector properties including sensor noise, optical distortion, depth of field, and FOV. These limit the attainable precision and resolution.
- Geometry of the system and scene, affecting view angles, triangulation angles, and the epipolar geometry. The measurement error field is usually not homogenous or isotropic within the FOV.
- Object radiometric properties including surface albedo, specular and diffuse surface reflection, sub-surface effects. Ideal materials exhibit Lambertian-like reflectance without sub-surface effects.
- Scene radiometric effects such as reflections. These are very dependent on the scene coding method used.

Owing to the number of parameters, it is much more difficult to quantify the accuracy and precision of a SL scanner compared to other measurement systems such as CMMs. A practical approach is to scan objects, which have previously been measured with more accurate methods. This methodology is widely used in metrology. In current literature, accuracy assessment is most often based on qualitative results [Güh00, SWPG05], and verifications on symmetrical geometries such as planes [SCR00], spheres and cylinders [BEHB95], or by comparison to e.g. a laser scan which is considered high quality [MT12].

A different approach to characterising measurement systems is by means of simulation. In [MDBS14] an SL simulation framework is presented, with physically based rendering of a scene under SL illumination, which focuses on benchmarking different coding strategies with respect to defocus, sub-surface effects and inter-reflections.

Spatial resolution has rarely been considered in literature. [GFS03] is an exception which aims to estimate a scanner's modulation transfer function, that is a characterisation of its response to spatial features of various frequencies. This is realised by scanning a sharp corner artefact, which naturally contains all frequencies.

6.1 Calibration Quality

As mentioned before, a proper camera model and calibration quality are crucial for accurate results in SL. Considering the calibration procedures described in Sec. 2.1, the following questions and issues arise:

- Which camera and lens model should be used for robust calibration? Some parameters may be correlated, and "confused" with each other, or lead to non-convergence [RVSS10].
- How many observations of the calibration board are necessary for convergence? I.e. how well are the camera model parameters observed in the images?
- At which angles should the calibration board ideally be observed? It is necessary to observe foreshortening to estimate focal lengths – however control point localisation becomes imprecise at shallow angles.

The choices made to accommodate these questions are usually based solely on simple statistics on the reprojection error, and the standard error of parameter estimates. Fig. 6.1 shows how the reprojection error and estimation errors are affected as the angular range of observation of a calibration board is increased. Apart from identifying point outliers, it is difficult to conclude the proper range from these plots. As such, the best evaluations of calibration quality are based on the quality of outcome, e.g. the correctness of point triangulation.

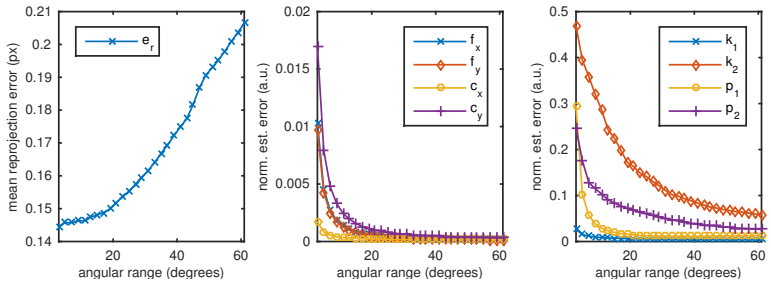


Figure 6.1: Mean reprojection error and normalised parameter uncertainties for the calibration of a SL system. In each calibration, the number of calibration board positions was 20, evenly distributed over the angular range.

6.2 Industry Standards for Accuracy

Stating accuracy figures is very important where SLs systems are used in industrial settings such as quality assurance and metrology. Hence, it is very interesting to know the exact performance of a system, and to be able to compare it to that of others. This need has spawned some efforts to develop industry standards for this purpose, but no established standard has thus far emerged. The International Organization for Standardization (ISO) standard 10360-2 is sometimes used, but was developed for optical probes attached to CMMs. ISO Committee 213 Working Group 10 is working on the formulation of a suitable standard for optical scanners in general. The standardisation organisation ASTM International has founded Committee E57 for the same purpose. The only currently published industry standard (as of late 2015) is the national German standard VDI 2634 [Gui13].

Figure 6.2: Artifacts for accuracy assessment according to VDI 2634(2). From [Contribution G].

The VDI 2634 standard demands to scan a dumbbell and a flat artifacts (see

Fig. 6.2) at several positions in the scanner's FOV. A number of error measures are determined by fitting spheres and planes to the data and determining worst case deviations. An example is shown in Fig. 6.3, where the error in dumbbell sphere distance, SD, is determined at seven positions in the FOV, for different angular ranges of calibration as described in Sec. 6.1. From these results, it can be concluded that a large angular range with much observed foreshortening yields slightly more accurate results.

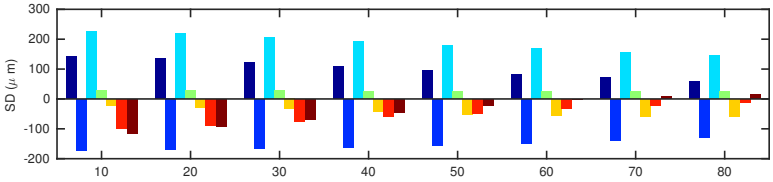


Figure 6.3: Sphere distance errors according to the VDI 2534 standard. The error is shown for seven different positions in the FOV and for different angular ranges used in the calibration procedure.

[Contribution G] provides a thorough study of the various calibration factors and their influence on the accuracy as quantified by means of VDI 2634.

6.3 System and Scene Geometry

The FOV of a SL system is given by the intersection of the projector and the camera(s) individual FOVs. As stated above, the accuracy of SL scanners can not be assumed isotropic within this FOV. Particularly, the triangulation angle θ varies, such that small amounts of projector light and camera pixel noise lead to different depth and lateral errors, as shown in Fig. 6.4.

Considering ideal image coordinates $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ in two views, and their real, observed coordinates $\tilde{\boldsymbol{q}}_1$ and $\tilde{\boldsymbol{q}}_2$, as illustrated in Fig. 6.4. With the assumption of parallel rays, and by simplifying such that $|\boldsymbol{q}_1\tilde{\boldsymbol{q}}_1| = |\boldsymbol{q}_2\tilde{\boldsymbol{q}}_2| = \epsilon$, the width (lateral error) and height (depth error) of the "error parallelogram" become

$$\epsilon_{\text{lateral}} = \sqrt{\frac{2\epsilon^2}{1+\cos\theta}} \qquad \epsilon_{\text{depth}} = \sqrt{\frac{2\epsilon^2}{1-\cos\theta}} \quad . \tag{6.1}$$

These errors dependency on θ are shown graphically in Fig. 6.5.

The geometric error also has a dependency on scene geometry in that shallow

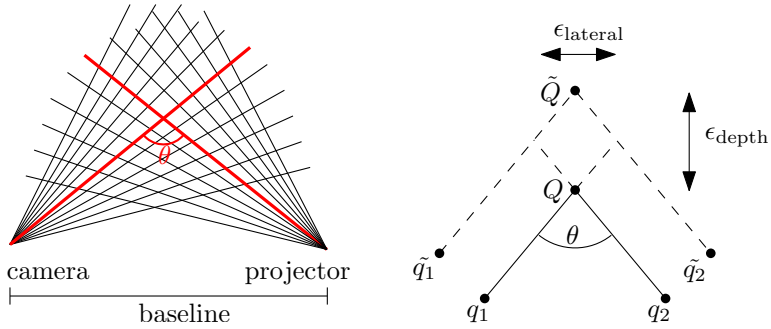


Figure 6.4: Left: 2-D illustration of the areas projecting into individual uncertainty regions in the camera and projector. Quadrilateral areas illustrate that both depth and lateral resolution vary within the system FOV. Right: considering the error in triangulation of a point with non-ideal observations.

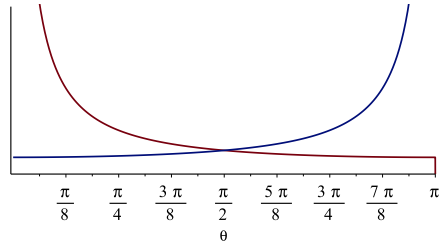


Figure 6.5: The lateral error $\epsilon_{\text{lateral}}$ (blue curve) and ϵ_{depth} (red curve), as defined by Eqn. 6.1, as a function of the triangulation angle θ.

illumination or viewing angles on the object surface make the decoding process less precise, since small errors in localisation lead to large 3-D errors.

6.4 Radiometric Effects

Many factors can lead to degraded performance in SL. In order to discuss this, it needs to be considered how light interacts with an object. In reality, the interaction between the projected light and the object is quite complex.

An example of a difficult to scan material is glass, as illustrated in Fig. 6.6. It is common to treat such surfaces with "dulling spray" containing small plastic particles which increase surface roughness and create a variety of surface normals at the microscopic levels. For even better reflectance properties, "chalk spray" can be used. It is seen that such treatment is strictly necessary for these objects to yield acceptable results. These procedures add small amounts of material to the object surface, however most often with effects much smaller than the error [Contribution H].

The sum of light can be thought of as a collection of rays or photons. These will penetrate a certain depth into the object, and undergo one or several of the following effects:

- Absorption: the photon energy is absorbed by the material.
- Transmission: the ray passes through the material.
- Reflection: at the interface of two materials, grains or other substructures, the path of the ray is redirected according to the law of reflection.
- Refraction: the path of the ray is redirected and follows the law of refraction.

From a macroscopic perspective the sum of these effects may be described as surface scattering, i.e. reflection into many directions due to surface roughness, subsurface scattering, which is single or multiple reflection/refraction events occurring at a distance below the surface and absorption of light.

With SL, only those rays are observed, which, after interacting with the object, reach the camera sensor. In a broader sense, this part of the light may be the result of multiple reflections, see Fig.6.7. Light which is reflected according to the law of reflection, is observed as a specular highlight, while other light reaching

Figure 6.6: SL scans of a household glass with no treatment (left), dulling spray (middle) and chalk spray (right). Top: photograph of the treated objects. Bottom: SL point clouds acquired with binary Gray coding.

the camera sensor may be diffusely reflected. For many materials, in particular metals, the specular reflection has many times higher irradiance than diffusely reflecting regions. In ideal diffuse reflection, the amount of light observed, or luminance, is independent of the view direction, which is known as Lambertian reflectance. In non-opaque material, part of the light will penetrate the object surface. On its way into and out of the material, refraction occurs according to Snell's law. Inside the material, it may be reflected one or multiple times. This is called sub-surface scattering. In homogenous half-infinite material, the probability of scattering of a single ray is characterised by the scattering coefficient.

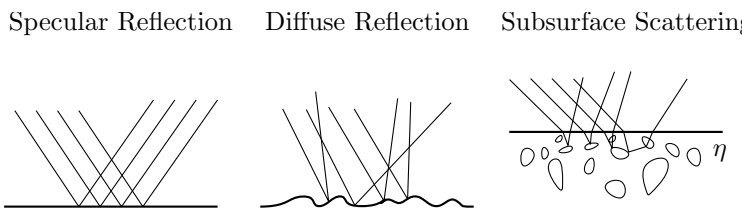


Figure 6.7: Optical interaction between illumination and the observed surface. Left: specular reflection at smooth surfaces. Middle: diffuse reflection due to surface roughness. Right: subsurface effects due to semi-transparency. Light rays undergo refraction at the boundary, then reflect once or multiple times.

It should be noted that sub-surface effects are wavelength dependent, with light penetrating deeper into the object at higher wavelengths. This is seen clearly in Fig. 6.8 which shows a sample of meat illuminated with a binary SL pattern at different wavelengths.

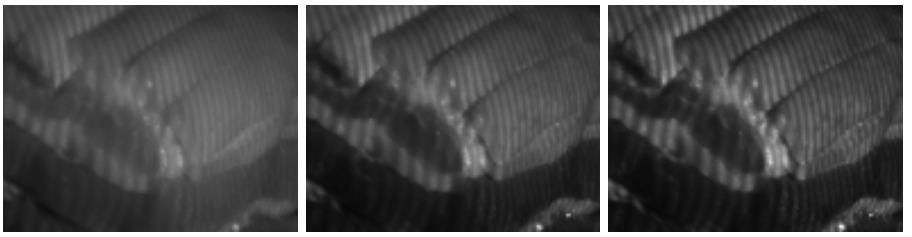


Figure 6.8: The wavelength-dependency of subsurface effects shown on a sample of meat illuminated with a binary SL pattern at different wavelengths. Left: 630 nm (red). Middle: 550 nm (green). Right: 450 nm (blue).

One of the most direct assumptions underlying accurate SL results is that light rays originating from the projector are reflected only at the object surface.

Thus, for SL, the ideal situation is one where reflection occurs only at the object surface [IKL+10]. In addition, due to the limited dynamic range of projectors and cameras, specular highlights are often too bright for successful decoding, so Lambertian like reflection without sub-surface effects is desired.

6.4.1 Direct and Global Light

Light captured by the camera in a SL system can be considered to be of two origins, the direct and global illumination. Direct illumination is due to reflection of the projector light. Global illumination is indirect, and due to reflection, sub-surface scattering and other effects. In SL global illumination is unwanted and may lead to artifacts, noise and bias [CSL08].

A separation of the scene light into direct and global components, or "descattering", can be done by illuminating single pixels, and observing the scene under such illumination as in [SMK05]. A much faster separation can be obtained by means of high frequency illumination, as shown by Nayar et al. [NKGR06]. This method relies on the observation that the global component can be considered invariant to spatial shifts of a high frequency illumination pattern. This allows for direct-global separation using e.g. binary lines or a shifted checkerboard pattern. It is also possible using high frequency sinusoidal patterns. Considering the N-shift PS sequence, as described in Sec. 2.3, the observed camera images are

$$I_n^c = A^c + B^c \cos\left(2\pi\left(\frac{n}{N} - \phi\right)\right) \quad , \tag{6.2}$$

where ϕ is the (unshifted) sinusoidal phase at that pixel coordinate. Since the global illumination can be considered constant under shift of the sinusoidal patterns, separation can be achieved with [NKGR06]:

$$L_{\text{global}} = 2A^c - B^c \quad L_{\text{direct}} = B^c \quad . \tag{6.3}$$

The same principle of separation holds for the phase at the principle frequency, ϕ. This means that direct-global separation is implicitly done with PS based methods if their illumination patterns are of high frequency. Figure 6.9 shows direct-global separation based on a PS sequence on a scene with concave regions. The "Modulated PS" method builds on this principle by modulating the patterns such that they are high-frequent both in both horizontal and vertical directions, which increases their separation power [CSL08]. In "Micro PS", high frequency patterns are used for both coding and phase-unwrapping, making both resistant to global effects [GN12].

The influence of direct and global light on SL results in biological tissue is investigated in [Contribution H].

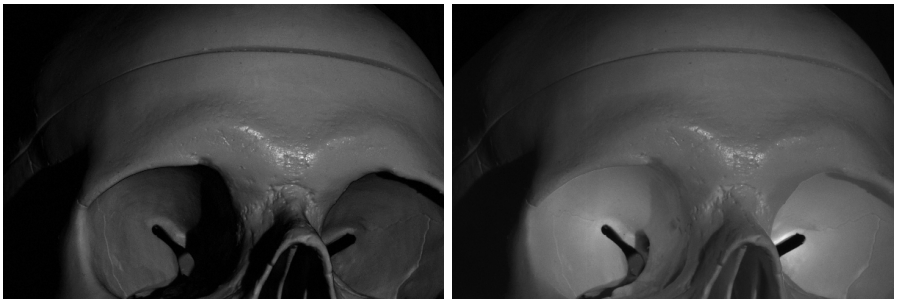


Figure 6.9: Separation of light into direct (left) and global (right) components by means of high frequency sinusoidal SL patterns. Clear differences are seen in the concave orbits. Some of the specular highlight was not separated entirely.

6.5 Contributions

As shown above, accuracy and precision have many influencing factors in practical SL systems. In order to investigate which accuracy can be obtained under ideal conditions, and which design parameters dictate this performance, a study was performed and is presented in [Contribution G]. Here the quality of calibration and scene coding is quantified by means of industry standard VDI 2634. This is one of few studies considering the whole scan process and quantifying accuracy and precision using an established standard. As such, it provides useful guidance to developers and users of SL systems.

I most practical scenarios, radiometric effects such as subsurface scattering play an important role for the quality of SL results. While recent literature has begun to address this, little work has been done to measure the error pertaining to these effects. [Contribution H] provides a quantification of bias and variance of SL scans of biological tissues; muscle, skin and fat, showing a clear deterministic component, and an error model which allows to remove this bias.

Chapter 7

Conclusion

SL is an optical measurement method of tremendous flexibility. It offers great advantages over many other methods of 3-D measurements for digitisation, interaction, tracking and metrology. Consequently, a large body of literature has been published on the subject, ranging from calibration to scene coding, applications and performance evaluation. This thesis has provided an overview of the literature and common methods used for SL.

Starting with the calibration of SL systems, the fundamental principles and previous approaches were presented in Ch. 2. A method was devised and described in [Contribution B] for the convenient and accurate calibration of such systems based on the current best methods. This method is implemented in [Contribution C] and used in the medical tracking device presented in [Contribution D].

Real time SL is a rather new possibility, and enables many new applications. The thesis has discussed the main challenges with real time SL in Ch. 3, and provided new and existing methods to overcome many of these challenges.

The problem of object tracking and pose estimation was considered in Ch. 4, and a methodology for fast tracking and filtering based on real time SL presented. This approach is used for motion tracking in MRI scanners in Ch. 5, and enables prospective motion correction in this setting.

Lastly, accuracy and precision parameters were considered for SL in Ch. 6. It was shown how both geometric and radiometric factors affect SL performance, which is important both for the medical application of this thesis, and many other future applications of SL.

Returning to the thesis objectives of Sec. 1.2, the first objective was to realise real time SL. This was accomplished by developing a hardware triggered setup,

fast OpenGL-based rendering of patterns and a multi-threaded software program to capture, decode, triangulate and visualise the result as shown in Ch. 3. The software implementation "SLStudio" is presented in [Contribution C], and freely available. The problem of motion artifacts in real-time SL was addressed with a correction scheme in [Contribution E].

Real-time pose tracking was the second objective. This problem was addressed by implementing fast projective ICP based tracking and by employing EKF to reduce noise in the tracking result. The correctness and accuracy of this approach was verified by means of marker based tracker which was coordinate aligned by means of hand-eye-calibration in Ch. 4. [Contribution A] relates to this approach by introducing the ICP based tracking method.

The objective of making a system suitable for head motion correction in medical scanners originally motivated this work. It was shown in Ch. 5 and [Contribution D,F] how SL can be made compatible with medical scanners, and point clouds are generated in the strong MRI magnetic field without disturbing the patient. An alternative approach could be based on passive stereo vision, however this is beyond the scope of this thesis. With 3-D active shape models for initialisation and subject-specific feature tracking, very accurate head tracking may be possible with a passive setup. This would simplify the hardware aspects, possibly allowing to place the sensors themselves in the MRI tunnel.

For an application such as head tracking, accuracy is paramount. Thus, a thesis objective was to investigate the factors influencing accuracy under ideal conditions. Many of these factors were discussed in Ch. 6. In [Contribution G], an investigation is made to determine which camera models, calibration routines and scene coding strategies provide the best results as quantified by an industry standard. An outlook to future work, which was outside the scope of this thesis is the determination of SL scanner performance by means of simulation with physics based rendering and Monte Carlo simulation. Such an approach of obtaining ground truth was used in [MDBS14] to benchmark scene coding methods. Similarly, it could be used to predict geometric measurement errors, similar to the "Virtual CMM" technique used in metrology.

Under most practical situations, and in the medical motion correction application in particular, radiometric effects influence measurement results as shown in Ch. 6. A study was therefore conducted on biological tissue to quantify this influence and determine its deterministic and stochastic components, contained in [Contribution H]. Future work on this problem could include the use of models for light transport to compensate for subsurface effects. This accounts to describing the observed measurement error using e.g. the model of Holroyd [HL11], allowing to compensate for its effect in a future measurement.

In summary, it can be concluded that the thesis objectives were met.

Bibliography

[ADP12] Hafeez Anwar, Irfanud Din, and Kang Park. Projector calibration for 3D scanning using virtual target images. *International Journal of Precision Engineering and Manufacturing*, 13(1):125–131, 2012.

[AL10] Henrik Andreasson and Achim J. Lilienthal. 6D scan registration using depth-interpolated local image features. *Robotics and Autonomous Systems*, 58(2):157–165, 2010.

[AMGC02] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[BEHB95] J. A. Beraldin, S.F. El-Hakim, and F. Blais. Performance evaluation of three active vision systems built at the national research council of canada. In *SPIE Optical 3-D Measurement techniques III*, pages 352–361, 1995.

[BK04] Julia F. Barrett and Nicholas Keat. Artifacts in CT: recognition and avoidance. *Radiographics*, 24(6):1679–91, 2004.

[BKW+08] Michael D. Breitenstein, Daniel Kuettel, Thibaut Weise, Luc van Gool, and Hanspeter Pfister. Real-time face pose estimation from single range images. In *IEEE Conference on Computer Vision and Pattern Recognition*, number 1, pages 1–8, 2008.

[BL95] Gerard Blais and Martin D. Levine. Registering multiview range data to create 3D computer objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):820–824, 1995.

[BM92] Paul J. Besl and Neil D. McKay. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[Bou] Jean-Yves Bouguet. *Camera Calibration Toolbox for Matlab*. http://www.vision.caltech.edu/bouguetj/calib_doc/.

[Bra00] G. Bradski. The OpenCV Library. *Dr Dobbs Journal of Software Tools*, 25:120–125, 2000.

[Bro71] Duane C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.

[CBLJ03] Richard E. Carson, Craig Barker, Jeih-San Liow, and Calvin A. Johnson. Design of a motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction for the HRRT. In *IEEE Nuclear Science Symposium Conference Record*, volume 5, pages 3281–3285, 2003.

[CBS00] Frank Chen, Gordon M. Brown, and Mumin Song. Overview of three-dimensional shape measurement using optical methods. *SPIE Optical Engineering*, 39(1), 2000.

[CM91] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. In *IEEE International Conference on Robotics and Automation*, volume CH2969, pages 2724–2729, 1991.

[CMR11] Vincent Couture, Nicolas Martin, and Sébastien Roy. Unstructured light scanning to overcome interreflections. In *IEEE International Conference on Computer Vision*, number August, pages 1895–1902, 2011.

[CMR14] Vincent Couture, Nicolas Martin, and Sébastien Roy. Unstructured light scanning robust to indirect illumination and depth discontinuities. *International Journal of Computer Vision*, 108(3):204–221, 2014.

[CSK05] Dmitry Chetverikov, Dmitry Stepanov, and Pavel Krsek. Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing*, 23(3):299–309, 2005.

[CSL08] Tongbo Chen, Hans-Peter Seidel, and Hendrik P.A. Lensch. Modulated phase-shifting for 3D scanning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[CXJS09] Xiaobo Chen, Juntong Xi, Ye Jin, and Jin Sun. Accurate calibration for a camera–projector measurement system based on structured light projection. *Optics and Lasers in Engineering*, 47:310–319, 2009.

[Dan99] Kostas Daniilidis. Hand-Eye Calibration Using Dual Quaternions. *The International Journal of Robotics Research*, 18(3):286–298, 1999.

[DK04] Benoit Desjardins and Ella A Kazerooni. ECG-Gated Cardiac CT. *American Journal of Roentgenology*, 182(4):993–1010, 2004.

[DRS09] Jamil Draréni, Sébastien Roy, and Peter Sturm. Projector calibration using a markerless plane. In *International Conference on Computer Vision Theory and Applications*, volume 2, pages 377–382, 2009.

[DZS+06] Christian Dold, Maxim Zaitsev, Oliver Speck, Evelyn A Firle, Jürgen Hennig, and Georgios Sakas. Advantages and limitations of prospective head motion compensation for MRI using an optical motion tracking device. *Academic Radiology*, 13(9):1093–103, sep 2006.

[EW99] Garry A. Einicke and Langford B. White. Robust extended Kalman filtering. *IEEE Transactions on Signal Processing*, 47(9):2596–2599, 1999.

[FAHB11] Christoph Forman, Murat Aksoy, Joachim Hornegger, and Roland Bammer. Self-encoded marker for optical prospective head motion correction in MRI. *Medical Image Analysis*, 15(5):708–719, 2011.

[FH09] G. Frankowski and R. Hainich. DLP-Based 3D Metrology by Structured Light or Projected Fringe Technology for Life Sciences and Industrial Metrology. In *Proceedings of SPIE*, volume 7210, 2009.

[Fit03] Andrew W. Fitzgibbon. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(13-14):1145–1153, 2003.

[FLM92] O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera Self-Calibration: Theory and Experiments. In *European Conference on Computer Vision*, 1992.

[FME+02] R.R. Fulton, S.R. Meikle, S. Eberl, J. Pfeiffer, and C.J. Constable. Correction for head movements in positron emission tomography using optical motion-tracking system. *IEEE Transactions on Nuclear Science*, 49:58–62, 2002.

[FSV04] David Fofi, Tadeusz Sliwa, and Yvon Voisin. A comparative survey of invisible structured light. In *SPIE 5303, Machine Vision Applications in Industrial Inspection XII*, 2004.

[FWGV11] Gabriele Fanelli, Thibaut Weise, Juergen Gall, and Luc Van Gool. Real time head pose estimation from consumer depth cameras. *Pattern Recognition*, 6835:101–110, 2011.

[GA98] J.S. Goddard and M.A. Abidi. Pose and motion estimation using dual quaternion-based extended Kalman filtering. In *SPIE Conference on Three-Dimensional Image Capture and Applications*, volume 3313, pages 189–200, 1998.

[GAVN12] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. A Practical Approach to 3D Scanning in the Presence of Interreflections, Subsurface Scattering and Defocus. *International Journal of Computer Vision*, 102(1-3):33–55, aug 2012.

[Gen96] Z.J. Geng. Rainbow three-dimensional camera: New concept of high-speed three-dimensional vision systems. *Optical Engineering*, 35(2):376–383, 1996.

[Gen11] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 160(2):128–160, 2011.

[GFS03] Michael Goesele, Christian Fuchs, and Hans-Peter Seidel. Accuracy of 3D range scanners by measurement of the slanted edge modulation transfer function. In *International Conference on 3-D Digital Imaging and Modeling*, 2003.

[GHL03] C. Guan, Laurence G. Hassebrook, and Daniel L. Lau. Composite structured light pattern for three-dimensional video. *Optics express*, 11(5):406–17, mar 2003.

[GHYC04] Hongwei Guo, Haitao He, Yingjie Yu, and Mingyi Chen. Least squares calibration method for fringe projection profilometry. *Optical Engineering*, 44(3), 2004.

[GN12] Mohit Gupta and Shree K. Nayar. Micro Phase Shifting. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 813–820, 2012.

[GPvdK+11] Ali Gholipour, Martin Polak, Andre van der Kouwe, Erez Nevo, and Simon K Warfield. Motion-robust MRI through real-time motion tracking and retrospective super-resolution volume reconstruction. In *IEEE Conference on Engineering in Medicine and Biology Society*, jan 2011.

[GR10] Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: Whither we are? *Optics and Lasers in Engineering*, 48(2):133–140, 2010.

[Güh00] Jens Gühring. Dense 3D surface acquisition by structured light using off-the-shelf components. In *SPIE Photonics West 2001*, pages 220—-231, 2000.

[Gui13] Gabriele Guidi. Metrological characterization of 3D imaging devices. In Fabio Remondino, Mark R. Shortis, Jürgen Beyerer, and Fernando Puente León, editors, *Proceedings of SPIE*, may 2013.

[GWH08] Wei Gao, Liang Wang, and Zhan-Yi Hu. Flexible Calibration of a Portable Structured Light System through Surface Plane. *Acta Automatica Sinica*, 34(11):1358–1362, 2008.

[Har94] Ri Hartley. Euclidean reconstruction from uncalibrated views. In *Applications of invariance in computer vision*, pages 235–256, 1994.

[HHJC99] Peisen S. Huang, Qingying Hu, Feng Jin, and Fu-Pen Chiang. Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering*, 38(6):1065, 1999.

[HL11] Michael Holroyd and Jason Lawrence. An analysis of using high-frequency sinusoidal illumination to measure the 3D shape of translucent objects. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2985–2991, 2011.

[HMW+12] Michael Herbst, Julian Maclaren, Matthias Weigel, Jan Korvink, Juergen Hennig, and Maxim Zaitsev. Prospective motion correction with continuous gradient updates in diffusion weighted imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 67(2):326–38, feb 2012.

[HRD+12] S. Holzer, R. B. Rusu, M. Dixon, S. Gedikli, and N. Navab. Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2684–2689. Ieee, oct 2012.

[HS93] J.M. Huntley and H. Saldner. Temporal phase-unwrapping algorithm for automated interferogram analysis. *Applied optics*, 32(17):3047–3052, 1993.

[HS97] Janne Heikkilä and Olli Silvén. A Four-step Camera Calibration Procedure with Implicit Image Correction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1112, 1997.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple view geometry*. Cambridge University Press, 2nd editio edition, 2003.

[HZC03] Peisen S. Huang, Chengping Zhang, and Fu-Pen Chiang. High-speed 3-D shape measurement based on digital fringe projection. *Optical Engineering*, 42(1):163, 2003.

[IKL+10] Ivo Ihrke, Kiriakos N. Kutulakos, Hendrik P.A. Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. *Computer Graphics Forum*, 29(8):2400–2426, 2010.

[Joh97] Andrew E. Johnson. *Spin-images: a representation for 3-D surface matching*. PhD thesis, Carnegie Mellon University, 1997.

[JV11] Bing Jian and Baba C. Vemuri. Robust Point Set Registration Using Gaussian Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645, 2011.

[Kal60] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[KE12] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, 12(12):1437–1454, 2012.

[KHCZ14] Nikolaus Karpinsky, Morgan Hoke, Vincent Chen, and Song Zhang. High-resolution, real-time three-dimensional shape measurement on graphics processing unit. *Optical Engineering*, 53(2), 2014.

[KMK07] Makoto Kimura, Masaaki Mochimaru, and Takeo Kanade. Projector calibration using arbitrary planes and calibrated camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, number 4, pages 7–8, 2007.

[KSM+14] Andre Kyme, Stephen Se, Steven Meikle, Georgios Angelis, Will Ryder, Kata Popovic, Dylan Yatigammana, and Roger Fulton. Markerless motion tracking of awake animals in positron emission tomography. *IEEE Transactions on Medical Imaging*, 33(11):2180–90, 2014.

[Kui02] Jack B. Kuipers. *Quaternions and Rotation Sequences: A Primer With Applications to Orbits, Aerospace, and Virtual Reality*. Princeton University Press, 2002.

[KV06] Thomas P. Koninckx and Luc Van Gool. Real-time range acquisition by adaptive structured light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):432–445, 2006.

[LC08] Jiarui Liao and Lilong Cai. A calibration method for uncoupling projector and camera of a structured light system. In *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pages 770–774, 2008.

[LGG+14] Yongjiu Liu, Hao Gao, Qingyi Gu, Tadayoshi Aoyama, Takeshi Takaki, and Idaku Ishii. A Fast 3-D Shape Measurement Method for Moving Object. In *International Conference on Progress in Informatics and Computing*, pages 219 – 223, 2014.

[LHG03] Jielin Li, Laurence G. Hassebrook, and Chun Guan. Optimized two-frequency phase-measuring-profilometry light-sensor temporal-noise sensitivity. *Journal of the Optical Society of America A*, 20(1):106–115, 2003.

[LM94] Feng Lu and Evangelos Milios. Robot pose estimation in unknown environments by matching 2D range scans. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 249–275, 1994.

[LRG+00] Marc Levoy, Szymon Rusinkiewicz, Matt Ginzton, Jeremy Ginsberg, Kari Pulli, David Koller, Sean Anderson, Jonathan Shade, Brian Curless, Lucas Pereira, James Davis, and Duane Fulk. The Digital Michelangelo Project: 3D Scanning of Large Statues. In *Conference on Computer Graphics and Interactive Techniques*, pages 131–144, 2000.

[LS04] Ricardo Legarda-Saenz. Accurate procedure for the calibration of a structured light system. *Optical Engineering*, 43(2):464, 2004.

[LSWW08] Zhongwei Li, Yusheng Shi, Congjun Wang, and Yuanyuan Wang. Accurate calibration method for a structured light system. *Optical Engineering*, 47(5), 2008.

[LSYX06] T. Li, E. Schreibmann, Y. Yang, and L. Xing. Motion correction for improved target localization with on-board cone-beam computed tomography. *Physics in Medicine and Biology*, 51(2):253–267, 2006.

[LWL+10a] Kai Liu, Yongchang Wang, Daniel L. Lau, Qi Hao, and Laurence G Hassebrook. Dual-frequency pattern scheme for high-speed 3-D shape measurement. *Optics Express*, 18(5), 2010.

[LWL+10b] Kai Liu, Yongchang Wang, Daniel L. Lau, Qi Hao, and Laurence G. Hassebrook. Gamma model and its analysis for phase measuring profilometry. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 27(3):553–62, 2010.

[LXYG13] Lei Lu, Jiangtao Xi, Yanguang Yu, and Qinghua Guo. New approach to improve the accuracy of 3-D shape measurement of moving object using phase shifting profilometry. *Optics Express*, 21(25):30610–30622, 2013.

[LZ09] Shuangyan Lei and Song Zhang. Flexible 3-D shape measurement using projector defocusing. *Optics letters*, 34(20):3080–3082, 2009.

[Mag09] Martin Magnusson. *The Three-Dimensional Normal-Distributions Transform — an Efficient Representation for Registration, Surface Analysis, and Loop Detection*. PhD thesis, Örebro University, 2009.

[Mas01] T. Masuda. Generation of geometric model by registration and integration of multiple range images. In *IEEE International Conference on 3-D Digital Imaging and Modeling*, pages 254–261, 2001.

[Mat] MathWorks. *MATLAB Computer Vision System Toolbox*. http://www.mathworks.com/help/vision/.

[MDBS14] Esdras Medeiros, Harish Doraiswamy, Matthew Berger, and Claudio T Silva. Using Physically Based Rendering to Benchmark Structured Light Scanners. 33(7), 2014.

[MHSZ13] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. Prospective motion correction in brain imaging: a review. *Magnetic Resonance in Medicine*, 69(3):621–36, mar 2013.

[MJA70] D.M. Meadows, W.O. Johnson, and J.B. Allen. Generation of surface contours by moiré patterns. *Applied Optics*, 9(4):942–947, 1970.

[MT12] Daniel Moreno and Gabriel Taubin. Simple, Accurate, and Robust Projector-Camera Calibration. In *International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 464–471, 2012.

[MWZ13] Leah Merner, Yajun Wang, and Song Zhang. Accurate calibration for 3D shape measurement system using a binary defocusing technique. *Optics and Lasers in Engineering*, 51(5):514–519, 2013.

[NDI+11] Richard A. Newcombe, Andrew J. Davison, Shahram Izadi, Pushmeet Kohli, Otmar Hilliges, Jamie Shotton, David Molyneaux, Steve Hodges, David Kim, and Andrew Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011.

[NG12] Shree K. Nayar and Mohit Gupta. Diffuse structured light. In *IEEE International Conference on Computational Photography*, pages 1–11, 2012.

[NKGR06] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics*, 25(3):935, 2006.

[OJP+10] Oline Vinter Olesen, Morten R. Jørgensen, Rasmus R. Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen. Structured Light 3D Tracking System for Measuring Motions in PET Brain Imaging. In *SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*, volume 7625, 2010.

[Ole11] Oline Vinter Olesen. *Markerless 3D Head Tracking for Motion Correction in High Resolution PET Brain Imaging*. PhD thesis, Technical University of Denmark, 2011.

[PA82] J.L. Posdamer and M.D. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17, 1982.

[PP10] Soon-Yong Yong Park and Go Gwang Park. Active calibration of camera-projector systems based on planar homography. In *International Conference on Pattern Recognition*, pages 320–323. Ieee, aug 2010.

[PT97] Yani Picard and Christopher J. Thompson. Motion correction of PET images using multiple acquisition frames. *IEEE Transactions on Medical Imaging*, 16(2):137–144, 1997.

[QTCF03] Chenggen Quan, Cho Jui Tay, Lujie Chen, and Yu Fu. Spatial-fringe-modulation-based quality map for phase unwrapping. *Applied Optics*, 42(35), dec 2003.

[RHHL02] Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. Real-time 3D model acquisition. *ACM Transactions on Graphics*, 21(3), 2002.

[RL01] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *Conference on 3-D Digital Imaging and Modeling*, pages 145–152. IEEE Comput. Soc, 2001.

[RRT97] Carsten Reich, Reinhold Ritter, and Jan Thesing. White light heterodyne principle for 3D-measurement. In *SPIE 3100, Sensors, Sensor Systems, and Sensor Data Processing*, volume 3100, pages 236–244, 1997.

[RVSS10] Carlos Ricolfe-Viala and Antonio-Jose Sanchez-Salmeron. Lens distortion models evaluation. *Applied optics*, 49(30):5914–28, oct 2010.

[SC01] Xianyu Su and Wenjing Chen. Fourier transform profilometry: A review. *Optics and Lasers in Engineering*, 35(5):263–284, 2001.

[SC08] Zhan Song and Ronald Chung. Use of LCD Panel for Calibrating Structured-Light-Based Range Sensing System. *IEEE Transactions on Instrumentation and Measurement*, 57(11):2623–2630, 2008.

[SCR00] G. Sansoni, M. Carocci, and R. Rodella. Calibration and performance evaluation of a 3-D imaging sensor based on the projection of structured light. *IEEE Transactions on Instrumentation and Measurement*, 49(3):628–636, jun 2000.

[SEH12] Mili Shah, Roger D. Eastman, and Tsai Hong. An overview of robot-sensor calibration methods for evaluation of perception systems. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, page 15, 2012.

[SFPL10] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 2010.

[SHK94] David A. Simon, Martial Hebert, and Takeo Kanade. Real-time 3-D pose estimation using a high-speed range sensor. *IEEE International Conference on Robotics and Automation*, pages 2235–2241, 1994.

[SLH84] V. Srinivasan, H.C. Liu, and M. Halioua. Automated phase-measuring profilometry of 3-D diffuse objects. *Applied Optics*, 23(18):3105–3108, 1984.

[SMFF07] Joaquim Salvi, Carles Matabosch, David Fofi, and Josep Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing*, 25(5):578–596, 2007.

[SMK05] Steven M. Seitz, Yasuyuki Matsushita, and Kiriakos N. Kutulakos. A theory of inverse light transport. In *IEEE International Conference on Computer Vision*, volume II, pages 1440–1447, 2005.

[SSR+12] Jessica Schulz, Thomas Siegert, Enrico Reimer, Christian Labadie, Julian Maclaren, Michael Herbst, Maxim Zaitsev, and Robert Turner. An embedded optical tracking system for motion-corrected magnetic resonance imaging at 7T. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 25(6):443–453, 2012.

[SWPG05] Filip Sadlo, Tim Weyrich, Ronald Peikert, and Markus Gross. A practical structured light acquisition system for point-based geometry and texture. In *Eurographics Symposium on Point-Based Graphics*, pages 89–145. Ieee, 2005.

[SXH+97] R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In *Proceedings SPIE*, volume 3100, pages 245–253, 1997.

[SZ10] Xianyu Su and Qican Zhang. Dynamic 3-D shape measurement method: A review. *Optics and Lasers in Engineering*, 48(2):191–204, 2010.

[SZRvH05] Jan-Jakob Sonke, Lambert Zijp, Peter Remeijer, and Marcel van Herk. Respiratory correlated cone beam CT. *Medical Physics*, 32(4):1176, 2005.

[TFR99] Emanuele Trucco, Andrea Fusiello, and Vito Roberto. Robust motion and correspondence of noisy 3-D point sets with missing data. *Pattern Recognition Letters*, 20(9):889–898, sep 1999.

[TIK82] Mitsuo Takeda, Hideki Ina, and Seiji Kobayashi. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *Journal of the Optical Society of America*, 72(1):156, jan 1982.

[TL89] Roger Y. Tsai and Reimar K. Lenz. A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration. *IEEE Transactions on Robotics and Automation*, 5(3), 1989.

[TL94] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *ACM Conference on Computer Graphics and Interactive Techniques*, pages 311–318, New York, New York, USA, 1994. ACM Press.

[TM83] Mitsuo Takeda and Kazuhiro Mutoh. Fourier transform profilometry for the automatic measurement of 3-D object shapes. *Applied Optics*, 22(24):3977, 1983.

[Tsa86] Roger Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[VaM98] R.J. Valkenburg and a.M. McIvor. Accurate 3D measurement using a structured light system. *Image and Vision Computing*, 16(2):99–110, feb 1998.

[WC91] Clarence Wust and David W. Capson. Surface profile measurement using color fringe projection. *Machine Vision and Applications*, 4:193–203, 1991.

[WFS11] Patrick Wissmann, Frank Forster, and Robert Schmitt. Fast and low-cost structured light pattern sequence projection. *Optics Express*, 19(24):24657, 2011.

[WLEZ13] Yajun Wang, Jacob I. Laughner, Igor R. Efimov, and Song Zhang. 3D absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique. *Optics express*, 21(5):6631–6636, 2013.

[WLL+10] Yongchang Wang, Kai Liu, Daniel L. Lau, Qi Hao, and Laurence G Hassebrook. Maximum SNR pattern strategy for phase shifting methods in structured light illumination. *Journal of the Optical Society of America A*, 27:1962–1971, 2010.

[WLV07] Thibaut Weise, Bastian Leibe, and Luc Van Gool. Fast 3D Scanning with Automatic Motion Compensation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[WOP+11] Jakob Wilm, Oline V. Olesen, Rasmus R. Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen. Real time surface registration for PET motion tracking. In *Scandinavian Conference on Image Analysis*, volume 6688 LNCS, pages 166–175, 2011.

[WSF11] Patrick Wissmann, Robert Schmitt, and Frank Forster. Fast and accurate 3D scanning using coded phase shifting and high speed pattern projection. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 108–115, 2011.

[WZO11] Yajun Wang, Song Zhang, and James H Oliver. 3D shape measurement technique for multiple rapidly moving objects. *Optics Express*, 19(9):8539–45, apr 2011.

[YPL+12] Yongkai Yin, Xiang Peng, Ameng Li, Xiaoli Liu, and Bruce Z. Gao. Calibration of fringe projection profilometry with bundle adjustment strategy. *Optics Letters*, 37(4):542, 2012.

[ZCG+12] Chao Zuo, Qian Chen, Guohua Gu, Shijie Feng, and Fangxiaoyu Feng. High-speed three-dimensional profilometry for multiply objects with complex shapes. *Optics Express*, 20(17):19493, 2012.

[ZCS02] Li Zhang, Brian Curless, and Steven M Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic

Programming. In *International Symposium on 3D Data Processing Visualization and Transmission*, pages 1–13, 2002.

[ZH06] Song Zhang and Peisen S. Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8):083601, aug 2006.

[Zha99] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *IEEE International Conference on Computer Vision*, volume 1, pages 666–673, 1999.

[Zha05] Song Zhang. *High-resolution, Real-time 3-D Shape Measurement.* PhD thesis, Stony Brook University, 2005.

[Zha10] Song Zhang. Recent progresses on real-time 3D shape measurement using digital fringe projection techniques. *Optics and Lasers in Engineering*, 48(2):149–158, feb 2010.

[ZLY07] Song Zhang, Xiaolin Li, and Shing-Tung Yau. Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction. *Applied Optics*, 46(1):50–57, 2007.

[ZS94] W.-S. Zhou and X.-Y. Su. A Direct Mapping Algorithm for Phase-measuring Profilometry. *Journal of Modern Optics*, 41(1):89–94, 1994.

[ZXYW14] Yueyi Zhang, Zhiwei Xiong, Zhe Yang, and Feng Wu. Real-time scalable depth sensing with hybrid structured light illumination. *IEEE Transactions on Image Processing*, 23(1):97–109, 2014.

Appendix A

Fast and Practical Head Tracking in Brain Imaging with Time-of-Flight Camera

Conference	IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Seoul Oct. 27 – Nov. 2 2013
Presentation	Poster Session
Published in	IEEE Nuclear Science Symposium Conference Record 2013
doi	10.1109/NSSMIC.2013.6829270

This contribution relates to the head tracking methodology described in Ch. 4 and the medical application detailed in Ch. 5. It represents the early state of our tracking approach, without Kalman Filtering, and using a commercially available TOF camera instead of the real time SL system which was developed subsequently. The commercial availability and low cost of TOF sensors makes them potential candidates for tracking. They suffer from general biases however, and as we explore in this study, their data quality cannot match that of our SL based system.

Fast and Practical Head Tracking in Brain Imaging with Time-of-Flight Camera

Jakob Wilm, Oline V. Olesen, Rasmus Ramsbøl Jensen, Liselotte Højgaard, and Rasmus Larsen

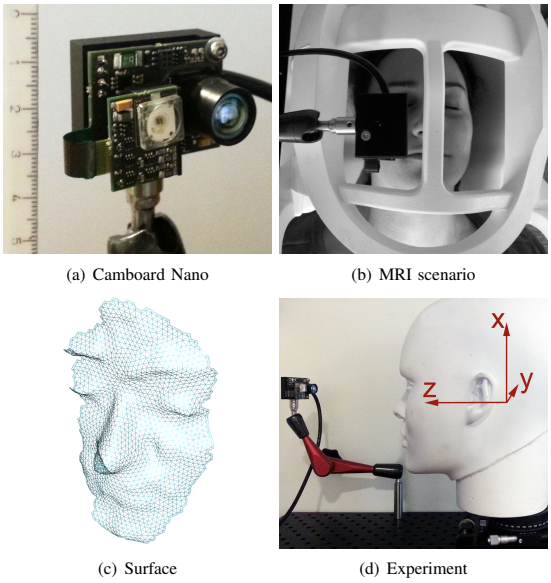
***Abstract*—This paper investigates the potential use of Time-of-Flight cameras (TOF) for motion correction in medical brain scans. TOF cameras have previously been used for tracking purposes, but recent progress in TOF technology has made it relevant for high speed optical tracking in high resolution medical scanners. Particularly in MRI and PET, the newest generation of TOF cameras could become a method of tracking small and large scale patient movement in a fast and user friendly way required in clinical environments.**

We present a novel methodology for fast tracking from TOF point clouds without the need of expensive triangulation and surface reconstruction. Tracking experiments with a motion controlled head phantom were performed with a translational tracking error below $2\,\mathrm{mm}$ and a rotational tracking error below $0.5°$.

I. Introduction

New opportunities for motion correction in medical imaging have emerged with recent improvements in computer vision. Several systems have been introduced for external motion tracking in medical scanners, but none of them are used on a daily basis in clinical settings with the exception of a few research sites. The main challenges still to be overcome are robustness and tracking accuracy and ease of operation for the medical staff. Patient throughput has become a very important issue nowadays and most scanning procedures do not allow for time consuming patient preparation or complicated system setups. As a consequence, current clinical implementations of motion correction methods are based on image data, which has the inherent disadvantage that intra-frame motion can not be corrected and assumptions must be made about the nature of the motion [1].

For these reasons, a single-vision-based motion correction method without markers is a very attractive prospect. However, the problem of reliably identifying and tracking natural features has so far led most authors to propose tracking systems with markers. Examples for MRI include multiple camera systems [1], [2] and single camera systems with advanced markers [3], [4]. Substituting the markers with structured light also provides accurate tracking as demonstrated for brain PET in [5] but structured light systems require multiple vision/viewing angles. Commercial tracking systems are also available (e.g. Northern Digital Inc., Claron Technology Inc., Metria Innovation Inc., etc.) which are branded to be universally applicable to intra-operative tracking and rigid body tracking in non-MRI environments.



(a) Camboard Nano (b) MRI scenario

(c) Surface (d) Experiment

Fig. 1. (a) Camboard Nano. (b) Camboard Nano mounted onto MRI head coil. (c) Triangulated surface reconstruction from TOF point cloud of the subject in (b). (d) Experimental setup of a head phantom mounted onto motion stages. Coordinate system definitions are shown for the ground truth (GT).

In this work we propose a markerless head tracking solution from a single vision by means of a time-of-flight (TOF) camera. This technology has seen significant improvements in terms of depth resolution and accuracy since the first commercial systems emerged in the early 2000's. Compared to other depth sensors (stereo vision, structured light), TOF cameras provide a much simpler, single-package monocular setup with fewer issues related to occlusion. Compared to laser range sensors, it provides faster frame rates and lower system costs. TOF cameras have also successfully been applied to medical imaging problems. Examples include respiratory motion detection [6], 3D endoscopy [7], and radiotherapeutic patient positioning [8]. However, to our knowledge, TOF cameras have not yet been applied to tracking for motion correction.

Currently, several manufacturers offer TOF cameras. These include pmdtechnologies, Fotonic, SoftKinetic, Panasonic,

Manuscript received December 10, 2013.

J. Wilm, O. V. Olesen, R. R. Jensen and R. Larsen are with the Department of Applied Mathematics and Computer Science, Technical University of Denmark (corresponding email: jakw@dtu.dk).

L. Højgaard is with the Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet, Copenhagen University Hospital, Denmark.

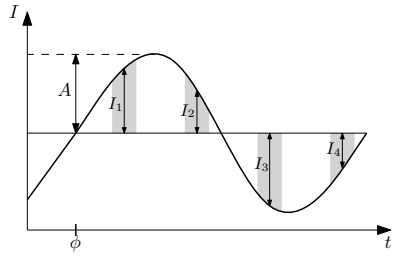


Fig. 2. Correlation signal at the TOF camera CMOS sensor. Gray bars indicate that the signal is integrated for a short time. φ denotes the phase shift, which is recovered from the sampled intensities and easily converted into a distance measure. Modified from [9].

MESA Imaging. This work is based on the Camboard Nano (Fig. 1(a)), a reference implementation of the pmd PhotonICs 19k-S3 TOF CMOS sensor. As of Feb. 2013, this is the smallest TOF camera on the market with a very broad field of view, which also makes it suitable for in-bore tracking applications such as MRI demonstrated in Fig. 1(b) and Fig. 1(c). Further, the high frame rate of 90 fps and the relative large resolution of 160×120 pixels makes this camera favorable for head tracking.

A. TOF Principles

TOF cameras work on the principle of measuring the time it takes photons from an active illumination source to the reflecting object and back to the camera. Different approaches exist to measure this timespan accurately, and the following section will provide a brief description of phase-detection of a modulated light source (the method employed in the Camboard Nano).

In this method, the active near infrared illumination source is amplitude modulated with a carrier frequency, f, in the mega- to gigahertz range. On the receiver side, the incoming light intensity is correlated with the emitted signal and integrated four separate times per frame on a CMOS active sensor for a short period ($\sim 500\,\mu$s). The samples are spaced equally in time at intervals of $\pi/2$ and given by (I_1, I_2, I_3, I_4), see Fig. 2. The phase shift is determined from these four samples in every pixel as

$$\varphi = \arctan \frac{I_1 - I_3}{I_2 - I_4} \quad ,$$

which can be converted into depth measurements as

$$D = \frac{c}{4\pi f}\varphi \quad ,$$

where c is the speed of light. A camera-like amplitude image can also be extracted as

$$A = \frac{1}{2}\sqrt{(I_1 - I_3)^2 + (I_2 - I_4)^2} \quad .$$

Given a value of f, the maximum unambiguous depth range of a modulated light source camera becomes $c/(2f)$. By phase-unwrapping methods, the depth-range can be extended at the risk of large errors where discontinuities occur in the depth map. Reversely, a lower modulation frequency improves depth resolution. The Camboard Nano is designed for whole body gesture control with a modulation frequency of 30 MHz and a FOV of $90° \times 68°$ and an unambiguous depth range of about 5 m. This is more than required by our application, however, slew of the illuminating LED deters from using a higher modulation frequency in this product.

Sources of Errors: While TOF cameras combine the advantages of active range sensors with the speed of camera systems at an affordable price, it is known that TOF cameras are subject to profound noise and error in the acquired range image [10]. There are three main sources temporal and spatial noises and an systematic depth dependent error, commonly referred to as "wiggling error".

Temporal noise is typically observed in every pixel, but can be mitigated by averaging several range images. As pointed out by [10] and [11], it is beneficial to perform averaging in the raw image data, which is also possible with the Camboard Nano. Temporal noise can partly be attributed to dark current in the CMOS, which is expected to increase as the camera sensor temperature rises. Spatial noise is typically reduced by convolving the acquired range image with a smoothing kernel or an edge-preserving bilateral filtering operation. The latter is implemented and available in the API of the Camboard Nano.

Finally, the "wiggling error" is an important source of error in TOF images. It is a spatial and depth dependent systematic deviation from true distance measures [9].

In addition to TOF specific errors, the conversion of TOF depth measurements to 3D points requires a model of ray projection to the TOF camera. Commonly the TOF camera is modeled as a pin-hole camera with the standard 5 parameter model of intrinsic parameters. When wide-angle lenses are used, a lens distortion model is usually required. Both the intrinsic parameters and distortion coefficients can be calibrated based on the image of amplitudes A.

II. Methods and Experiments

In order to reduce the wiggling error, depth calibration was performed by capturing depth maps of planar surfaces at distances of 10 – 25 cm. The precision of plane positioning was $10\,\mu$m. Wiggling error was then considered to be linear in all pixels. Prior to the modeling the captured planes were filtered and rotated perpendicular to the image axis using principal component analysis. Fig. 3 shows the image plane before and after correction at a distance of 144 mm. Point clouds of a plan were captured at ten arbitrary positions within the tracking depth range of 10–25 cm for evaluation of the correction.

The TOF camera was calibrated with the Camera Calibration Toolbox for Matlab by Jean-Yves Bouguet[1]. It returns camera parameters according to the pin-hole model and radial and tangential distortion parameters. With our TOF camera, this method leads to reprojection errors below 1 pixel.

Our TOF tracking algorithm estimates the relative pose of the head by means of the iterative closest point (ICP) algorithm. Considering the problem of aligning one TOF point cloud $\mathbf{P} = \{p_i\}$, to another, $\mathbf{Q} = \{q_j\}$, where individual point

[1] http://www.vision.caltech.edu/bouguetj/calib_doc/

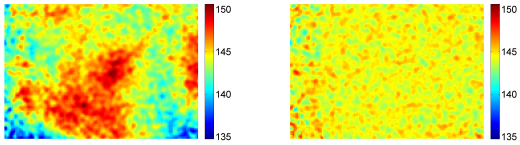


Fig. 3. The TOF image plane before (left) and after (right) depth calibration at a distance of 144 mm (where color represents the depth in mm). The standard deviation (SD) of the depth before and after correction are 2.3 mm and 0.74 mm respectively.

correspondences are not known. In order to address the issue of missing point correspondence, the ICP algorithm iteratively matches points in $\mathbf{P}$ to their nearest neighbors in $\mathbf{Q}$ to form the pairs $(p_i, \hat{q}_i)$ for $i = 1 \dots N_{\text{pairs}}$, it minimizes an error metric, and then transforms $\mathbf{P}$ using the result. Our error metric is

$$\sum_{i=1}^{N_{\text{pairs}}} \|\mathbf{T}_{\text{TOF}}\, p_i - \hat{q}_i\|^2 \quad , \tag{1}$$

where $\mathbf{T}_{\text{TOF}}$ is a homogenous 4×4 matrix representing the rigid transformation, while p_i and q_i are homogenous point vectors.

Our pose estimation procedure exploits the fact that the point clouds are generated by projecting out from depth measurements that lie in a regular grid. We use this spatial arrangement to identify points that lie on the boundary of surfaces. Such points are characterized by the fact that there are empty entries in their 8-neighborhood in the regular grid. Consequently, a single loop through all points allows us to identify such points. This then makes robust alignment with the ICP algorithm possible. By discarding point matches to boundary points, convergence is achieved despite of partial overlap of the surfaces. In scattered point clouds, such determination of boundary points usually would require surface reconstruction or other complex operations on the point clouds.

To evaluate our tracking performance with the TOF camera, we employ an experimental setup as shown in Fig. 1(d). In this setup, a phantom of a head is mounted onto a system of two micro-translation stages and one rotation stage, which allows us to perform exact transformations of the head pose ($0.01\,\text{mm}$ and 0.01^o respectively). There spatial transformations will be the basis of our evaluation, and we refer to them as "ground truth" (GT). The TOF camera is mounted at a distance of approximately $15\,\text{cm}$ from the phantom head and offset slightly ($1 - 2\,\text{cm}$) from the mid-sagittal plane in the starting position. Independently, a series of translational (50 mm) and rotational (15^o) movements were performed by the phantom, and recovered by means of ICP alignment of the TOF point clouds. The integration time of the TOF camera was set to a constant value of 400 μs and the raw image data were averaged over 4 samples.

To compare the GT with the results of the ICP alignment, data from both has to be brought into a common coordinate system. We do this by constructing a series of homogenous transformation matrices, $\mathbf{T}_{\text{GT},i}$, $\mathbf{T}_{\text{TOF},i}$, $i = 1 \dots N_{\text{poses}}$ for the GT and the results of the ICP alignment respectively. Rotation angles on the rotation stage are considered described

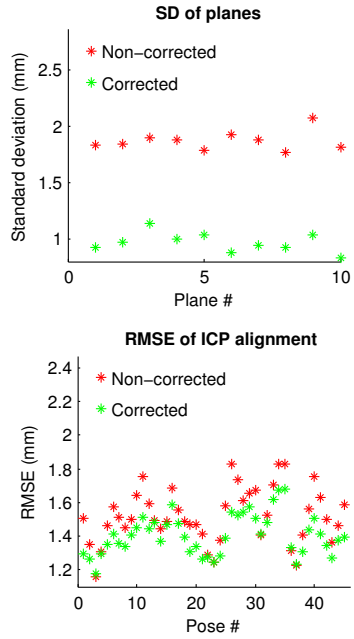


Fig. 4. The effect of "wiggling" correction. Left: SD of a plan at ten arbitrary positions. Right: RMSE of ICP alignments of phantom data with and without correction.

by Euler angle ψ_{GT} around the first coordinate axis as shown in Fig. 1(d). We then solve for the best rigid transformation, $\mathbf{H}$, bringing the two data sets into alignment by means of a closed form fitting procedure introduced by Shah [12]

$$\underset{H}{\text{argmin}} \sum_{i=1}^{N_{\text{poses}}} \|\mathbf{H}\, \mathbf{T}_{\text{TOF},i} - \mathbf{T}_{\text{GT},i}\| \quad .$$

We then extract and compare corresponding translation vector components and Euler angles of the two transformation series.

III. Results

Fig. 4 shows the effect of correcting for the wiggling error both on a series of planes and on the root mean square error (RMSE) of the ICP alignments of the phantom in different poses. The standard deviation (SD) of the point clouds to a fitted plane was improved from a mean of $1.8\,\text{mm}$ to a mean of $0.97\,\text{mm}$ after "wiggling error" correction. Clearly, the alignment RMSE was also decreased significantly ($p < 10^1 1$) as a result of the correction.

Fig. 5 shows the results of the tracking experiment. The RMSE of translations ($[T_x, T_y, T_z]$) over all poses are $[1.88,\ \ 1.83,\ \ 1.10]\,\text{mm}$ and the RMSE of rotations are $[0.42,\ 0.16,\ 0.47]^\circ$ for rotations around x, y, and z (ψ, θ, ϕ) respectively. A correlation is observed between the translational errors and the residual rotations. This is seen as an overall bias where the rotation is slightly underestimated and the translation is slightly overestimated. This suggests that

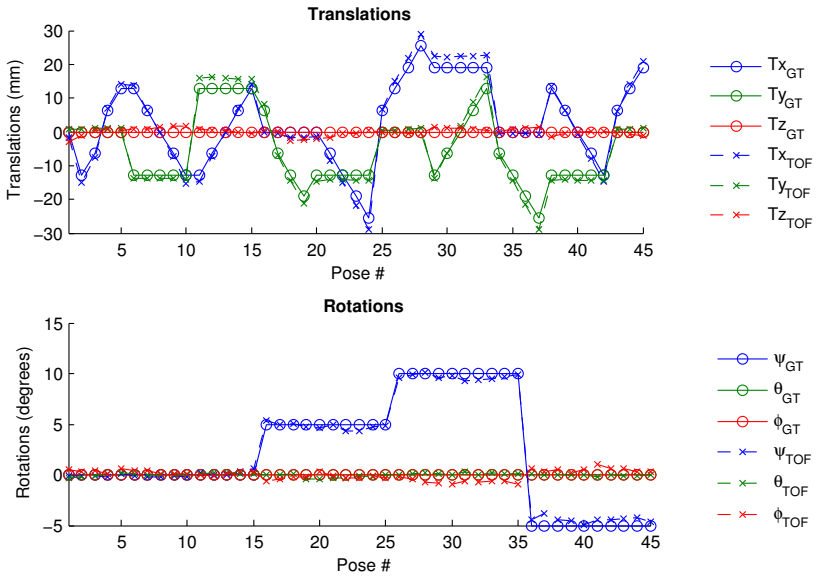


Fig. 5. Tracking results of the phantom experiment. Ground truth (GT) motion is compared to our method using the TOF camera for six motion parameters: three translations (top) and three rotations (bottom).

tracking results could potentially be improved by different coordinate alignment.

IV. Discussion and Conclusions

We have performed an investigation to determine the usefulness of the newest generation of TOF cameras for rigid body tracking, in particular 3D head tracking for motion correction in medical imaging. A processing pipeline was presented that corrects for the most detrimental error in TOF imaging and utilizes the regular spatial arrangement of depth samples. The tracking precision achieved in our experiments suggests that head pose tracking with TOF cameras has become feasible in cases where it is medium to large motion that needs to be resolved. It was shown that using appropriate correction and filtering of the TOF data, system accuracy can be improved significantly, and as a consequence, large scale movement on the order of millimeters can be detected. More specifically, our experiments demonstrated that tracking RMS errors below $2\,\mathrm{mm}$ for translational motion and below 0.5° for rotational motion can be achieved in close range application with the Camboard Nano TOF camera.

As such, we believe that TOF cameras are suitable today for patient monitoring in imaging procedures, and possibly for motion correction in high resolution brain imaging in the future. We emphasize that the major advantages of the presented TOF tracking approach is the simple and compact design with a single vision preferable for small geometries with limited sight just like the bore holes of a medical scanner. Furthermore, our system does not rely on markers attached to the subject with failure prone and time consuming marker attachment.

References

[1] Zaitsev, M., Dold, C., Sakas, G., Hennig, J., Speck, O.: Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. NeuroImage **31**(3) (2006) 1038–1050

[2] Qin, L., van Gelderen, P., Derbyshire, J.A., Jin, F., Lee, J., de Zwart, J.A., Tao, Y., Duyn, J.H.: Prospective head-movement correction for high-resolution MRI using an in-bore optical tracking system. MR in Med **62**(4) (2009) 924–934

[3] Rotenberg, D., Chiew, M., Ranieri, S., Tam, F., Chopra, R., Graham, S.J.: Real-time correction by optical tracking with integrated geometric distortion correction for reducing motion artifacts in functional MRI. MR in Med **69** (2013) 734–748

[4] Andrews-Shigaki, B.C., Armstrong, B.S.R., Zaitsev, M., Ernst, T.: Prospective motion correction for magnetic resonance spectroscopy using single camera Retro-Grate reflector optical tracking. J of MRI **33**(2) (2011) 498–504

[5] O.V., Sullvian, J., Morris, E., Mulnix, T., Paulsen, R., Højgaard, L., Roed, B., Larsen, R.: List-mode PET motion correction using markerless head tracking: proof-of-concept in human studies. IEEE TMI **32** (2013) 200–209

[6] Penne, J., Schaller, C., Hornegger, J., Kuwert, T.: Robust real-time 3D respiratory motion detection using time-of-flight cameras. Intl J of CARS **3** (2008) 427–431

[7] Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B., Hornegger, J.: Time-of-Flight 3-D Endoscopy. LNCS **5761** (2009) 467–474

[8] Placht, S., Stancanello, J., Schaller, C., Balda, M., Angelopoulou, E.: Fast time-of-flight camera based surface registration for radiotherapy patient positioning. Medical Physics **39**(1) (2012) 4–17

[9] Kahlmann, T., Remondino, F., Ingensand, H.: Calibration for increased accuracy of the range imaging camera Swissranger. Proc of IEVM (2006)

[10] Rapp, H.: Experimental and theoretical investigation of correlating TOF-camera systems. Diploma Thesis, U of Heidelberg (2007) 59–63

[11] Lindner, M., Schiller, I., Kolb, A., Koch, R.: Time-of-Flight sensor calibration for accurate range sensing. CVIU **114**(12) (2010) 1318–1328

[12] Shah, M.: Comparing two sets of corresponding six degree of freedom data. CVIU **115**(10) (2011) 1355–1362

Appendix B

Accurate and Simple Calibration of DLP Projector Systems

Conference	SPIE Photonics West, San Francisco 1. – 6. February 2014
Presentation	Oral Presentation
Published in	Proceedings SPIE 8979, Emerging Digital Micromirror Device Based Systems and Applications VI, 897909
doi	10.1117/12.2038687

The efficient and accurate calibration of single-camera SL systems is treated in this contribution (see Sec. 2.1 for an introduction to the topic). The approach follows the concept of Zhang and Huang [ZH06] of letting the projector "indirectly" capture images of a calibration target and calibrate like a camera. The main idea is to use RBFs as robust and accurate local models to translate the feature coordinates from camera to projector space. We also provide a solution to index diamond pixel arrays found in some DLP projectors.

Accurate and Simple Calibration of DLP Projector Systems

Jakob Wilm[1,2] and Oline V. Olesen[1,2] and Rasmus Larsen[1]

[1]Department of Applied Mathematics and Computer Science,
Technical University of Denmark,
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark;

[2]Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet,
Copenhagen University Hospital, Denmark;

ABSTRACT

Much work has been devoted to the calibration of optical cameras, and accurate and simple methods are now available which require only a small number of calibration targets. The problem of obtaining these parameters for light projectors has not been studied as extensively and most current methods require a camera and involve feature extraction from a known projected pattern. In this work we present a novel calibration technique for DLP Projector systems based on phase shifting profilometry projection onto a printed calibration target. In contrast to most current methods, the one presented here does not rely on an initial camera calibration, and so does not carry over the error into projector calibration. A radial interpolation scheme is used to convert features coordinates into projector space, thereby allowing for a very accurate procedure. This allows for highly accurate determination of parameters including lens distortion. Our implementation acquires printed planar calibration scenes in less than 1 s. This makes our method both fast and convenient. We evaluate our method in terms of reprojection errors and structured light image reconstruction quality.

Keywords: Projector calibration, radial basis functions, structured-light, phase shifting profilometry

1. INTRODUCTION

While camera calibration has been intensively studied, and state of the art methods have evolved, the calibration of projectors and camera projector systems is usually done using non-optimal methods, which sacrifice accuracy for speed or convenience. This is despite the fact that accurate determination of the optical parameters is of utmost importance in many applications within optical surface profilometry, stereolithography or visual display.

Current methods for projector calibration are usually relying on calibrated sensors to observe projection output and determine the appropriate parameters. This approach has the inherent disadvantage of propagating the sensor calibration error into the estimates of the projector parameters. A notable exception is the work of Moreno et al.,[1] which shares some properties with our proposed method.

While we do use a camera to determine world-projector relationships, we do not rely on the camera being calibrated. As such, no calibration error is propagated.

The pinhole-model is well established for digital cameras, and due to their optical similarities to most projectors, also applicable here. The model is usually further extended with distortion parameters to account for effects that don't follow projective geometry.

Our lens model contains ten parameters – horizontal and vertical focal lengths, f_x, f_y, principal point coordinates c_x, c_y, skew coefficient α_c and lens distortion parameters $k_1 \dots k_5$. This is very similar to the widely accepted model of Heikkilä and Silvén,[2] and is the same as implemented in the popular camera calibration toolbox of Bouget[3] and in the OpenCV library.[4] Considering a world point in the camera's or projector's reference frame, $\boldsymbol{Q} = [x_w, y_w, z_w]^\top$, the pin hole camera projection gives the normalized image coordinates as

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} x_w/z_w \\ y_w/z_w \end{bmatrix}$$

Corresponding author: Jakob Wilm, E-mail jakw@dtu.dk, Telephone +45 45253716.

Spherical aberration gives rise to a radial distortion model according to

$$s(r) = 1 + k_1 r^2 + k_2 r^4 + k_5 r^6 \quad ,$$

with $r^2 = x^2 + y^2$, while the effects of imperfect lens alignment are termed tangential distortion, and modelled as

$$\boldsymbol{t}(x, y) = \begin{bmatrix} 2k_3 xy + k_4 \left(r^2 + 2x^2\right) \\ k_3 \left(r^2 + 2y^2\right) + 2k_4 xy \end{bmatrix} \quad .$$

This distortion effect is also called *decentering* or *thin prism* distortion, and was initially described by Brown.[5] The corrected normalised image coordinates of the point $\begin{bmatrix} u_n \\ v_n \end{bmatrix}$ with radial and tangential distortion are given by

$$\begin{bmatrix} u_c \\ v_c \end{bmatrix} = s\Big(\sqrt{u_n^2 + v_n^2}\Big) \begin{bmatrix} u_n \\ v_n \end{bmatrix} + \boldsymbol{t}(u_n, v_n) \quad .$$

Finally, the pixel coordinates of said point are given by

$$\boldsymbol{q} = \begin{bmatrix} u_p \\ v_p \end{bmatrix} = \begin{bmatrix} f_x & \alpha_c f_x & c_x \\ 0 & f_y & c_y \end{bmatrix} \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix}$$

Our method was developed for a new in-bore medical 3D scanner, which relies on accurate measurements, and requires a convenient calibration procedure that can be carried out by non-technical personal. To fulfil these goals, we have implemented a graphical user interface with visual feedback to ensure good coverage of the projector FOV and detection of feature point locations. This allows us to acquire a view of a calibration target in under one second, allowing for handheld operation of the projector or camera.

2. METHODS

Our method relies on fast acquisition of "projection sequences" with an industrial camera.

As mentioned earlier, our method does not rely on a calibrated camera. Instead we merely observe projection output in the presence of a planar checkerboard pattern. The projector encodes its pixels by means of *phase shifting profilometry* (PSP), and this allows us to establish world to projector plane correspondences.

We use a light grey checkerboard as the calibration target, which represents a trade off between robustness in corner detection and projector coordinate coding.

In recent iterations of their DLP technology, Texas Instruments has introduces DMDs with diagonal or *diamond* pixel arrays. The DLPC possesses a natural indexing of mirrors in this geometry, which has unique coordinates for every horizontal row and every other vertical column. Since we are interested in a linear mapping of pixel indices to position, we handle the diamond pixel array by assigning unique indices to every row and every column. With this definition, every other coordinate does indexes a "pseudopixel" – and in turn we encode only every other coordinate. See figure 2 for an illustration of the indexing convention for the DMD employed in our projector. It should be noted, that with this definition, calibration parameters and error estimates are expressed in units of our "pseudopixel", which in physical dimensions corresponds to roughly half of a mirror diagonal.

After determining world coordinates in the projector coordinate system, we calibrate it like a camera using the method of Zhang.[6]

3. PHASE SHIFTING STRUCTURED LIGHT

Our encoding of projector coordinates into the projected output is realised with the gray level *phase shifting profilometry* (PSP) method pioneered by Hung.[7] While many methods exist for codification in structured light,[8] PSP allows for dense scene encoding with a low number of patterns.

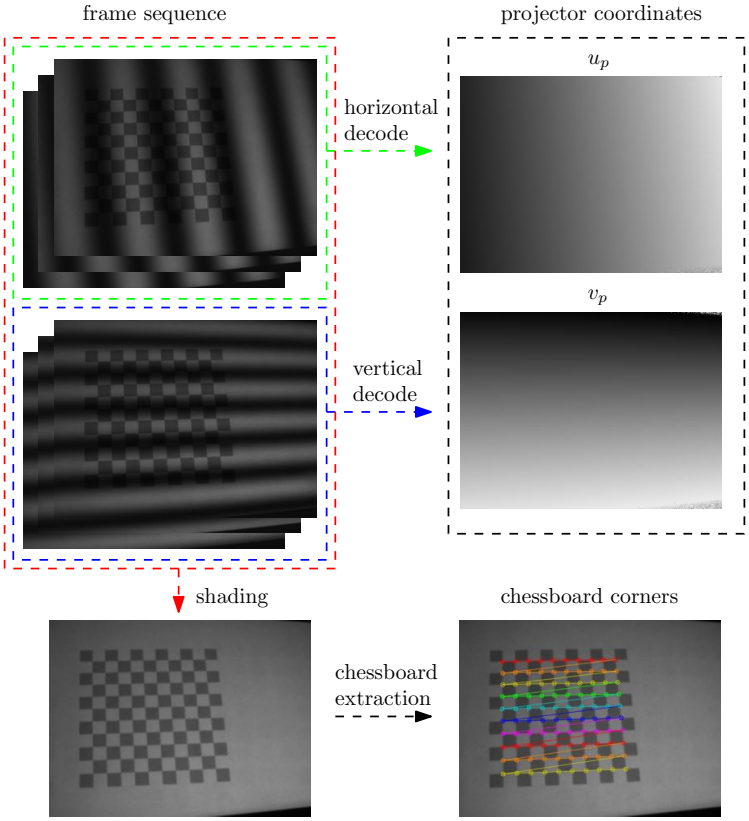


Figure 1. Flow diagram for the calibration procedure on one frame sequence. Horizontal and vertical encoding patterns are decoded to yield u_p and v_p maps. The patterns are also used to extract a general shading or intensity image, from which the chessboard corners are automatically extracted. These represent the checkerboard corners in the camera reference frame. By means of Gaussian radial basis functions, each set of corner coordinates is translated into projector coordinates.

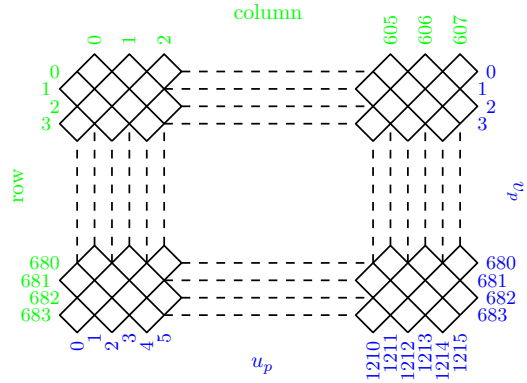


Figure 2. Diamond pixel DMD geometry for the DLP3000. The indexable screen coordinates are shown in green on top and left, while our pseudopixel coordinates are shown in blue on bottom and right. Note that in our pixel space, every other coordinate does not index a pixel, however, we do not need a dense pixel array for all subsequent steps.

Projector pixels are encoded first by their vertical (u_p) and then by their horizontal (v_p) coordinate. In the case of vertical encoding, the normalized pattern intensity may be described as

$$I_n^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\left(\frac{n}{3} - v_p\right)\right) \quad ,$$

where n indicates the pattern index, $n \in 1\ldots3$. The intensity of the n'th pattern as observed in the camera can be expressed as

$$I_n^c(u_c, v_c) = A^c + B^c\cos\left(\frac{2\pi n}{3} - \theta\right) \quad ,$$

in which θ denotes the phase of the sinusoidal sequence at that particular point. A^c is the intensity of the scene including ambient contributions, and B^c the intensity of reflected projector light. In terms of Fourier analysis, A^c can be considered the magnitude of the constant component, while B^c is the magnitude at the principle frequence, and θ its phase. As such, θ can easily be extracted by means of the fast fourier transform and scaled to the number of projector columns. The process for row codification is the same.

In order to reduce quantification error and noise, the number of encoded phases in PSP can be increased. The resulting map of θ then needs to be unwrapped in order to remove the resulting phase ambiguity. A reliable way of unwrapping is using a single phase *phase cue* sequence. We employ this strategy of using six patterns for the horizontal and six patterns for the vertical encoding with 8 phases. Thus, the total number of projected patterns is 12 per calibration target pose. This gives us for every frame sequence the encoded projector row and column, and by calculating B^c, a general intensity or *shading* image, which we use for chessboard extraction (see Fig. 1).

4. GAUSSIAN RADIAL BASIS FUNCTIONS

After horizontal and vertical frame sequence decoding and chessboard extraction, we seek to translate the chessboard corner coordinates from camera coordinates $\boldsymbol{q}_c$ into projector coordinates $\boldsymbol{q}_p$. Collecting such coordinates from a number of chessboard corners from multiple views allows us to calibrate the projector by treating it as a camera and calibrate it such.

In our case, we wish to interpolate at the points determined from chessboard detection, e.g. at a sub pixel level. While the underlying function is expected to be close to linear, due to the effects of lens distortion, a more complex function underlies the observed data.

We treat the translation of every newly extracted chessboard corner as an interpolation problem. While the projector coordinates were estimated at regular sites, we do not employ gridded data interpolation. The reason

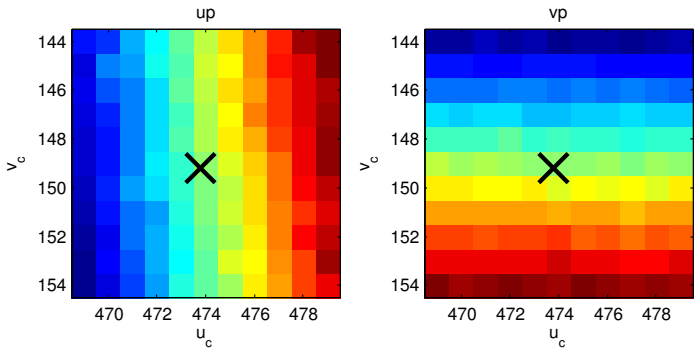


Figure 3. The interpolation problem exemplified on one chessboard corner location (black crosses). Left: the window of decoded projector columns. Right: same for projector rows.

is twofold; first we remove data by thresholding on the shading image, so we need to take care of missing data. Secondly, we wish to benefit from the advantages of using *radial basis function* (RBF) interpolators. Regular data interpolation is certainly the faster option, but in our application the complexity of RBFs does not become prohibitive.

In the RBF framework, a function is associated with each data point, and a regular set of equations solved to determine how each observation is explained as a linear combination of these functions. In order to ease our later analysis, we add a linear polynomial term $P(\boldsymbol{x}_i)$, such that the RBF will effectively describe the deviation from this global linear trend. To account for noisy data, we also add a regularization parameter k to our interpolator:

$$f_i = \sum_j \lambda_j \psi \|\boldsymbol{x}_i - \boldsymbol{x_j}\| + P(\boldsymbol{x}_i) + k\lambda_i \tag{1}$$

Common choices for $\psi(r)$ include a Gaussian function, $\psi(r) = \exp(-\alpha r^2)$ and the socalled Thin Plate Spline, $\psi(r) = r^2 \log(r)$. The choice of distance function influences the characteristics of the interpolator. Therefor we choose an RBF with a Gaussian distance function so as to give the highest weight to close neighbours while being robust against noise in the projector coordinate estimates. An influence region of around 100 neighbors seems reasonable, and correspondingly, as our interpolation sites are expressed in pixel coordinates, we have found good results with $\alpha = 1/5$.

All observations within a window around one detected chessboard corner are used to build the RBF interpolator. In practice this is done by forming the following equation from 1:

$$\begin{bmatrix} \boldsymbol{\Psi} + k\boldsymbol{I} & \boldsymbol{P}^\top \\ \boldsymbol{P} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{0} \end{bmatrix} \quad ,$$

where Ψ contains the basis functions, $\boldsymbol{P}$ linear polynomial terms and k is the scalar parameter which controls regularization – i.e. the balance between the exact RBF interpolator and the simple linear polynomium, in this case a plane.

As mentioned, the underlying function is expected to have a linear trend, as without the effects of lens distortion, there is a direct perspective relationship between camera and projector coordinates.

Since we are operating with a regularized regression model, we can define a *hat* or *smoother* matrix $\hat{\boldsymbol{H}}$ according to

$$\hat{\boldsymbol{H}} \begin{bmatrix} \boldsymbol{\Psi} + k\boldsymbol{I} & \boldsymbol{P}^\top \\ \boldsymbol{P} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{P}^\top \end{bmatrix} \quad ,$$

Projector – TI LightCrafter				Camera – Point Grey FL3-FW-03S1M-C				
DMD Chip	Exposure	Bit Depth	Gamma	Image Sensor	Exposure	Bit Depth	Gamma	Gain
TI DLP3000	16.667 ms	8 bit	1.0	Sony ICX618 CCD	16.667 ms	8 bit	1.0	0 dB

Table 1. Data and default parameters for our DLP projector and camera during calibration.

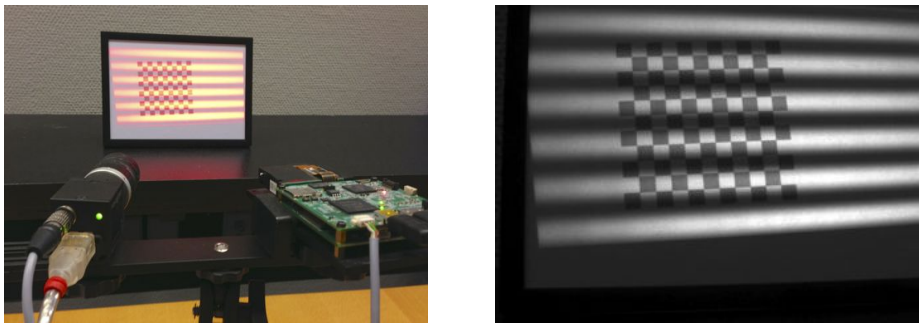


Figure 4. Left: Physical setup for our projector calibration method with camera (left) and projector (right). The calibration target is shown in the background. Right: corresponding frame captured by the camera.

and use the $\mathrm{Tr}(\hat{\boldsymbol{H}})$ as a measure of the *effective* degrees of freedom[9] of our model. We therefor choose k such that the effective number of degrees of freedom is 10, i.e. a little more than the 3 defining a plane, but much less than the 100 data points underlying each model.

5. IMPLEMENTATION

Our calibration setup is shown in figure 4. As seen, we have mounted the camera and projector in a fixed configuration, which allows us to also perform stereo calibration and triangulation world point coordinates using the structured light principle. Our calibration target is the checkerboard printed on a standard commercial laser printer using PostScript and spray glued onto the glass of a picture frame. This yields a highly accurate calibration target at a fraction of the cost of professionally manufactured offerings.

We use a LightCrafter DLP projector (Texas Instruments Inc.) with DLP3000 DMD (7.62 mm diagonal) and the FireWire industrial camera (Point Grey Research Inc.) FL3-FW-03S1M-C mono with Sony ICX618 CCD (4.00 mm diagonal). Table 1 shows default acquisition parameters. Theses values are chosen as a trade off between acquisition time and depth of field.

We have implemented our method as part of a complete structured light program in C++ utilizing the Qt, OpenCV and PCL (Point Cloud Library) libraries. See figure 5 for screenshots of the user interface. To enable fast pattern projection, we create a full screen OpenGL rendering context and upload the necessary patterns to the GPU as textures, which are rendered as screen filling quads. The camera interface uses libdc1394[10] to capture images from our FireWire industrial camera. In order to achieve a linear intensity response, we configure the projector to not use the typical spatiotemporal algorithms for visual image enhancement, and our camera to not perform gamma correction or signal amplification.

For each calibration pose, the corresponding series of captured frames is stored in the form of a frame sequence. Upon request, and when at least three frame sequences have been acquired, calibration is performed.

Checkerboard corners are automatically extracted using the appropriate functions in OpenCV's *calib3d* package. This returns the desired image coordinates with subpixel accuracy. If checkerboard extraction is unsuccessful, the frame sequence is excluded from further processing. Otherwise, the frame sequence is decoded to yield projector coordinates and calibrate the projector like a camera using OpenCV's built in functions.

Accurate capture of DLP projector output requires careful synchronisation of camera and projector exposure. While this is commonly realised using a trigger circuit, we opt for manual "software triggering", i.e. image projection and image capture is initiated through software commands. To account for frame buffering in the GPU and the projector and latency in camera triggering, we impose a short pause in between frames.

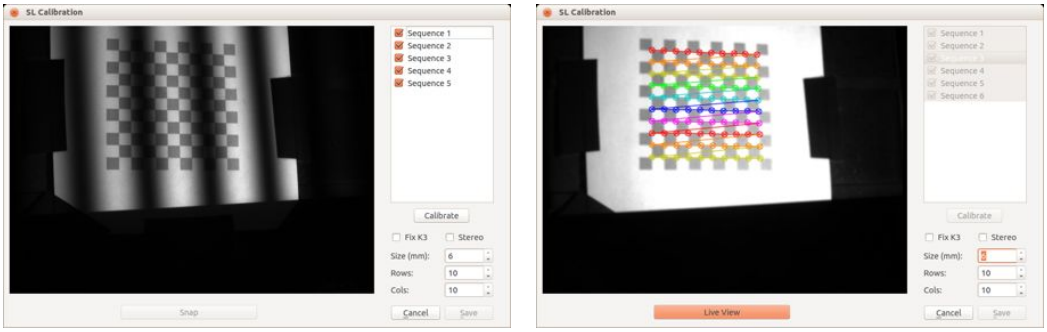


Figure 5. Video 1 Screenshots of our graphical user interface for calibration. Left: during calibration sequence acquisition. Right: during parameter estimation. http://dx.doi.org/doi.number.goes.here

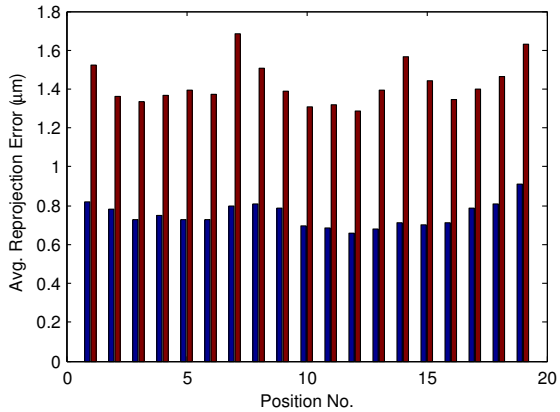


Figure 6. Average reprojection errors per pose for a calibration involving 18 calibration target poses for the camera (blue) and projector (red) respectively, measured in units of camera pixels and projector "pseudopixels" respectively.

6. RESULTS

Figure 6 shows reprojection error in a calibration involving 18 positions for the calibration target.

The individual reprojection errors are shown for the camera and for the projector on Fig. 7.

As the calibration target was moved through the intersecting FOVs of camera and projector three times, a positional dependence is aparent. This might indicate that the 5 parameter lens distortion model is not completely adequate. Low reprojection errors do not necessarily indicate a faithful calibration, as overfitting might explain low values. In this calibration case, a total of $100 \times 18 \times 2 = 3600$ constraints are given for a rather low parametric model. No outliers are visible from the individual error plots in Fig. 7.

To measure the quality of 3D point reconstruction with our calibration method and correction for lens distortion, we have aquired structured light images of a spherical bowling ball that has been painted to a diffuse surface. The 2x3 PSP pattern strategy was used. We have then fitted a sphere to the resulting point cloud and assessed the deviations and the structure of the errors in a histogram, see Figure 8.

As can be seen, errors increase towards the edges, as the angle between camera axis and normal increases. The RMS of errors over the foreground is 0.57 mm. The error distribution is dominated mainly by triangulation error at the edges of the bowling ball.

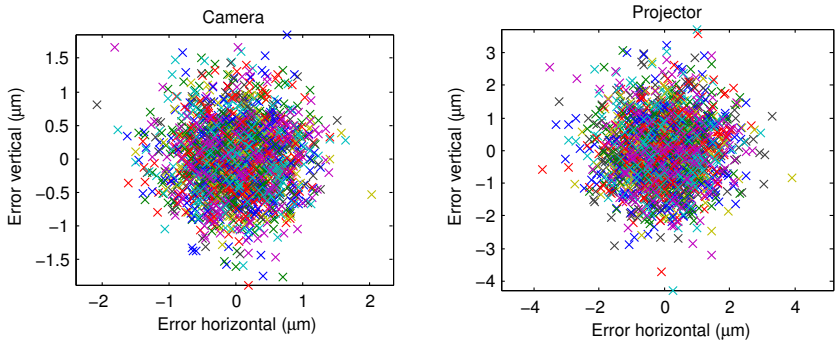


Figure 7. Individual reprojection errors with color coding corresponding to individual poses. Left: camera. Right: projector.

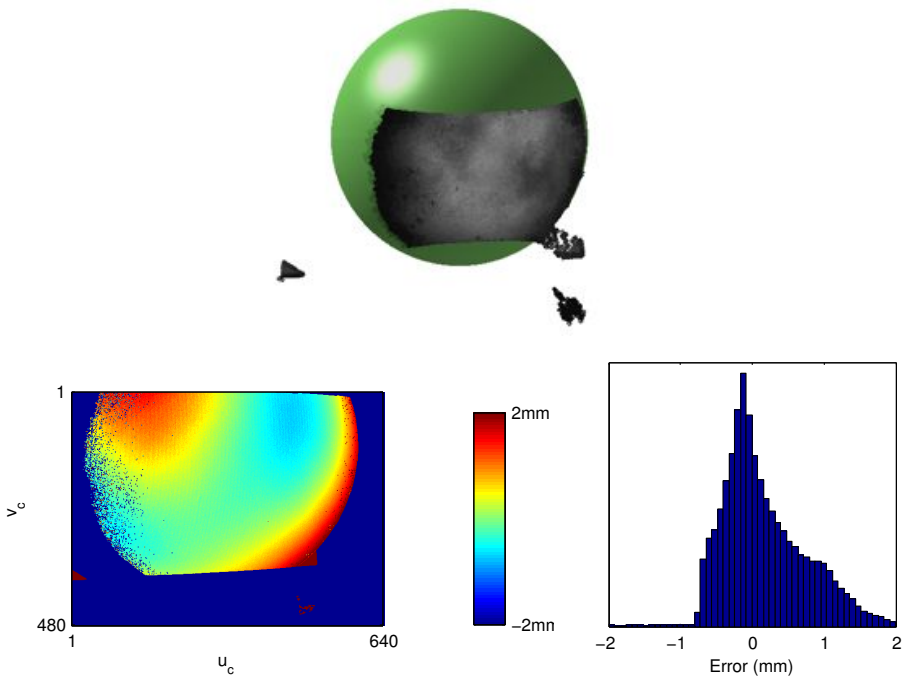


Figure 8. Structured light scan of a bowling ball. Top: sphere fitted to the triangulated data. Left: Plot of the error. Right: histogram of errors within the foreground. The RMS of errors is 0.57 mm.

7. CONCLUSIONS

We have shown a novel method and implementation of a convenient and accurate calibration method for light projectors. We utilize phase shifting profilometry for projector coordinate coding. Calibration points are translated into projector coordinates by means of Gaussian RBFs, which offers excelent trade-offs between accuracy and robustness. Visual feedback and direct camera/projector control enables operators to calibrate a lightprojector in a few minutes.

In addition, we determine projector lens distortion parameters with high accuracy and precision. Our method calibrates a camera at the same setting, and can perform a stereo calibration between projector and camera, which is particularly useful for single camera structured light applications.

Whilst most current methods require intricate setups or lengthy calibration procedures, we are able to calibrate very quick and accurately with only readily available equipment.

REFERENCES

[1] Moreno, D. and Taubin, G., "Simple, Accurate, and Robust Projector-Camera Calibration," *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission* , 464–471 (October 2012).

[2] Heikkila, J. and Silven, O., "Calibration procedure for short focal length off-the-shelf CCD cameras," *Proceedings of 13th ICPR* **1** (1996).

[3] Bouguet, J.-Y., "Camera Calibration Toolbox for Matlab." `http://www.vision.caltech.edu/bouguetj/calib_doc/`

[4] Itseez, "OpenCV, Open Source Computer Vision." `http://opencv.org/`

[5] Brown, D. C., "Decentering Distortion of Lenses," *Photometric Engineering* **32**, 444–462 (1966).

[6] Zhang, Z., "Flexible camera calibration by viewing a plane from unknown orientations," *The Proceedings of the ICCV 1999* (1999).

[7] Hung, D. D., "3D scene modelling by sinusoid encoded illumination," *Image and Vision Computing* **11**, 251–256 (June 1993).

[8] Salvi, J., Pagès, J., and Batlle, J., "Pattern codification strategies in structured light systems," *Pattern Recognition* **37**, 827–849 (April 2004).

[9] Hastie, T., Tibshirani, R., and Friedman, J., "The elements of statistical learning data mining, inference, and prediction," *Springer-Verlag* (2009).

[10] Douxchamps, D., "libdc1394: The API for IEEE1394." `http://damien.douxchamps.net/ieee1394/libdc1394/`

Appendix C

SLStudio: Open-Source Framework for Real-Time Structured Light

Conference	4th International Conference on Image Processing Theory, Tools and Applications (IPTA), 14. – 17. Oct. 2014
Presentation	Oral Presentation
Published in	Proceedings IPTA 2014
doi	10.1109/IPTA.2014.7002001

In this paper, we present our software platform for real time SL as described in Ch. 3. The software suite is generic and includes an implementation of the calibration method of [Contribution B], and a number of different scene coding methods. This is the first implementation capable of operating in real time and it is released as open source. As such, the contribution will be useful both for users with new applications of real time SL as well as research into new scene coding methods.

SLStudio: Open-Source Framework for Real-Time Structured Light

Jakob Wilm[1], Oline V. Olesen[1] and Rasmus Larsen[1]
[1] Department of Applied Mathematics and Computer Science,
Technical University of Denmark,
Matematiktorvet, Building 321, DK-2800 Kgs. Lyngby, Denmark
e-mail: jakw@dtu.dk
[2] Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet,
Copenhagen University Hospital, Denmark

***Abstract*— An open-source framework for real-time structured light is presented. It is called "SLStudio", and enables real-time capture of metric depth images. The framework is modular, and extensible to support new algorithms for scene encoding/decoding, triangulation, and aquisition hardware. It is the aim that this software makes real-time 3D scene capture more widely accessible and serves as a foundation for new structured light scanners operating in real-time, e.g. 20 depth images per second and more. The use cases for such scanners are plentyfull, however due to the computational constraints, all public implementations so far are limited to offline processing. With "SLStudio", we are making a platform available which enables researchers from many different fields to build application specific real time 3D scanners.**

The software is hosted at
`http://compute.dtu.dk/~jakw/slstudio`.

***Keywords*— Images acquisition systems and information extractionm, Image processing tools, Image and video processing, computer vision**

I. Introduction

The advent of inexpensive consumer depth sensors has lead to much interesting research in the fields of human computer interaction, 3D scene understanding, pose estimation and others. These new devices are used for a plethora of applications, ranging from scene capture and understanding to object digitization, motion tracking, human-computer interaction, and many others.

Current devices do however present different trade-offs between spatial and temporal resolution and system geometry, cost, and features. Many new applications call for a flexible device, which can be custom built and optimized for the problem at hand.

Currently, few software packages are publicly available for structured light, and to our knowledge, this is the first open-source software with real-time performance. The hardware components involved are rather inexpensive, such that one can implement a megapixel resolution, 20 Hz, device for under 2000 USD.

The software framework presented here is highly customizable to work with cameras and projectors of different manufacturers, and supports a palette of scene encoding/decoding strategies. The software, referred to as "SLStudio" is available open-source, encouraging active community engagement.

We developed SLStudio for motion tracking and correction in medical imaging [4], but the software has wide applicability, including and extending those applications which currently use off-the-shelve devices such as Microsoft Kinect, ASUS Xtion and others.

In recent years, a number of real time dense scene aquisition methods have emerged. *Time of Flight* work on the principle of measuring the time delay between emission and receival of light pulses or oscillations. While these sensors provide fast update rates, they suffer from high noise and bias [2], making them unsuitable for applications that require high accuracy.

Triangulation based methods require two viewpoints of a scene. In the passive case, these would consitute two cameras, and depth calculation is based on estimating disparity between the views. These methods most often require long processing time, however real-time implementations do exist, see e.g. [7] and [9]. The accuracy of these methods is limited, and especially weak in homogenous image regions.

Active triagulation methods employ laser light or projected patterns in order to add texture to the scene, and improve the accuracy and robustness of triangulation. The structured light concept is especially versatile in that different projected pattern sequences provide trade-offs between accuracy, robustness and update frequency. Single-shot methods employ a static pattern, and rely on neighborhood information for decoding, hence limiting their resolution. In order to acchieve real-time performance, however, emission and capture of light has to be accurately synchronized, and the further processing steps need to be optimized. For this reason, only relatively few systems have been presented in the literature. These include the works of Zhang [8] and Liu et al. [3]. To our knowledge, none of these real time structured light software implementations are publically available or open source.

II. Structured light principle

The fundamental principle underlying structured light methods is triangulation of corresponding points for a projector-camera pair. This is shown schematically in figure 1. A single camera is sufficient for scene acquisition, in that the projector acts as a second viewpoint. The distinct advantage over two-camera stereo is that texture is added by means of the projector light, which aids in the determination of correspondences.

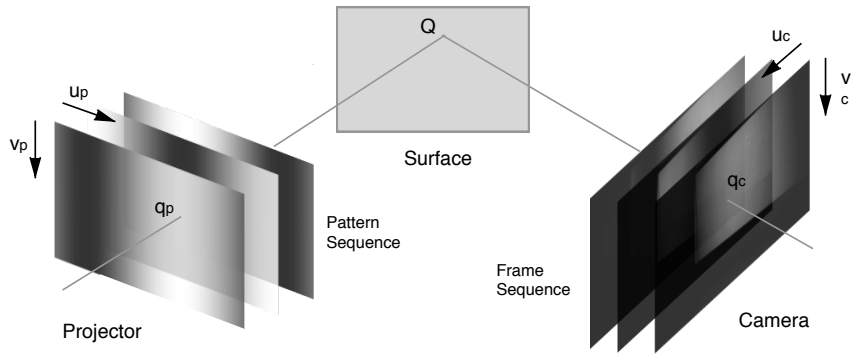


Fig. 1. Principle of structured light scanning with one camera and one projector, as implemented in SLStudio. Object surface points are determined by triangulating corresponding points in the camera and projector coordinate spaces.

Having control over the projected output pattern allows for direct coding of projector coordinates into a series of camera frames. Using an appropriate decoding method, projector coordinates can then be determined in the camera's view. Finally, the triangulation principle is employed to extract spatial coordinates, yielding 2.5D depth maps, possibly with texture information.

A calibration is usually performed prior to scanning, as this relates pixel coordinates to real world distances, and allows for metric reconstructions of the scene. In the calibration procedure, the optical properties of the projector and camera are determined, usually aided by a calibration target such as a black/white checkerboard with known spatial dimensions.

The steps involved in obtaining a single depth map are the following:

1) Scene encoding in which one or both projector coordinates are encoded into a *sequence of patterns*, which are then projected onto the scene in succession. Pattern strategies include *Phase Shifting Profilometry* (PSP), binary Gray codes, color coding, and many others. A review of such methods is given in [1].
2) Frame capture. One camera frame is acquired for each projected pattern. We refer to one set of camera frames as a *frame sequence*.
3) Decoding the frame sequence by means of an algorithm that matches and reverses the action of the pattern encoder. This determines for every camera pixel the corresponding projector coordinates. From this step, an intensity or shading image can usually be calculated, which provides color or gray-scale texture for point cloud rendering.
4) Triangulation, which is the process of turning the decoded projector coordinates into a metric 2.5D surface represented by points. Geometrically, this step corresponds to triangulation of points by intersecting corresponding rays originating from camera and projector. If only one projector coordinate was coded into the frame sequence, the intersection may be found as the solution to a linear set of three equations. With two coordinates encoded, the system is overdetermined, and can be solved using different triangulation algorithms. This reconstruction of points may be followed by further processing, i.e. *surface reconstruction*, in which a mesh representation is computed.

III. Implementation

"SLStudio" is written in C++ and makes use of OpenCV and Point Cloud Library for most of its computations, while the Qt framework is used for the graphical user interface, and multi-core processing. Figure 2 shows screenshots from within the running program. The main program window contains a 3D point cloud display and a histogram, to aid in avoiding over- or underexposure, and additionally, the decoded u_p map and texture information.

The software is comprised of the following modules, all implemented as abstract C++ classes, with a number of concrete implementations:

Calibrator is a generic interface to calibration methods. A number of patterns are provided, and the corresponding camera frames are then used to determine the intrinsics of the camera and projector lenses, and the geometric relation between the two. Point cloud are expressed in the camera's coordinate system after calibration. We provide one calibration method, which is highly accurate and very convenient. It captures frame sequences of a planar checker board using PSP and uses radial basis functions to translate world point coordinates into projector space. Projector lens distortion is determined and corrected for. Details are found in [6].

Projector models the interface for pattern projection. It features a direct projection mode in which the pattern is provided directly as a block of memory, and a deferred mode, in which patterns may be buffered for later display. The latter is convenient in the default *OpenGL* implementation, which allows patterns to be uploaded to GPU texture memory for fast display with vertical synchronization. This implementation has custom, compile-time determined versions for Linux X.org,

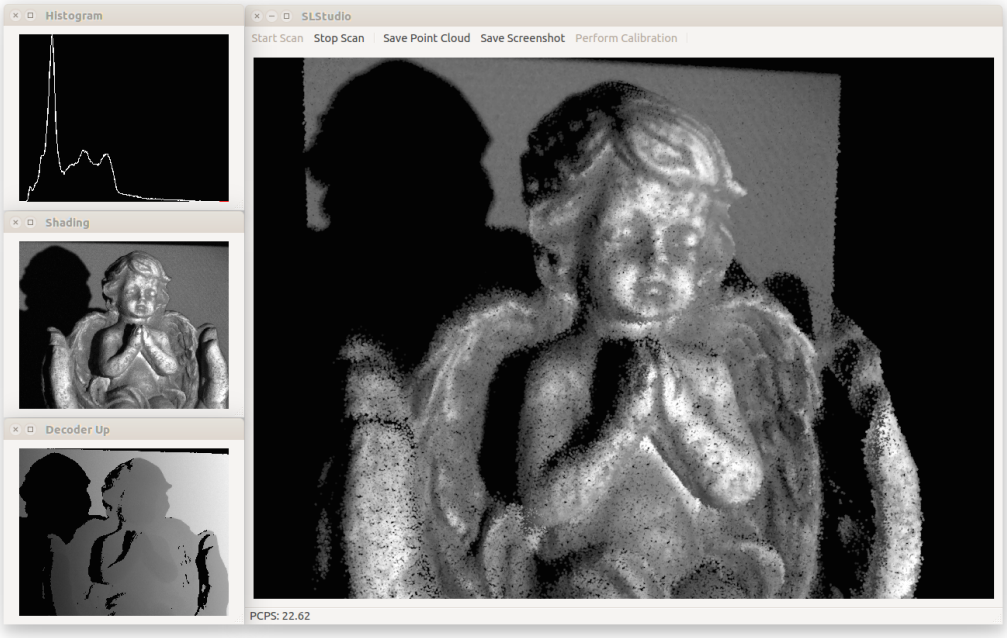


Fig. 2. Screenshot of SLStudio while acquiring 3D point clouds, showing camera histogram, decoded phase, shading information and the main window showing an angel figurine captured at 20 point clouds per second (640×512 pixels).

Windows and Mac OS X, to create a second fullscreen context, that does not interfere with the main screen, which displays the control GUI.

The **Camera** class implements communication with a camera. Two distinct modes are supported; software trigger and hardware trigger. In many industrial camera, hardware triggering enables fast frame acquisition by avoiding the overhead of initiating exposure through a software API, and having projector and camera exposure synchronized perfectly. Attention must be paid to exposure times – for instance, in DLP (Digital Light Projection, Texas Instruments Inc.), projected grayscale values will only be faithfully represented if the camera exposure is a multiple of the single pattern exposure time. For hardware triggering, a short electrical trigger pulse is sent to the camera at the start of each projection. For commercial projectors, a compatible trigger source is the analogue VSYNC signal found in VGA and DVI cables. The class is implemented as an object "factory", to support multiple cameras through the same interface. Concrete implementations exist for the camera APIs of IDS Imaging, Point Grey Research Inc., XIMEA, and the libdc1394 library for most industrial Firewire cameras.

A **Codec** is comprised of matching *Encoder* and *Decoder* classes, which have concrete implementations for different pattern strategies. A distinction is made between coding in one or two directions. While one is sufficient for triangulation, two directions are used where high accuracy is mandatory. The output of any decoding operation is termed u_p or v_p map, as it represent for every camera pixel coordinate the corresponding horizontal or vertical coordinate in the projector output. Implementations include *Phase Shifting Profilometry* (PSP) with and without phase unwrapping and binary *Gray coding*. The highest point cloud update frequency is acchieved with 3-step PSP, which requires only three patterns for scene encoding.

The **Triangulator** is the class involved in triangulation of points using predetermined camera and projector parameters and the u_p and/or u_p map obtained from decoding a frame sequence. We have implemented fast triangulation based on u_p or v_p using a pre-computed *determinant-tensor*[5], and algebraic triangulation. SLStudio represents point clouds internally in the PCL format, which can be written to disk or used for further processing.

Our software is modular and easily extensible to fit the specific requirements for a depth measurement device. The preferences dialog, shown in figure 3 shows the configurable options currently available.

The program employs task-based multithreading to share work among the CPU cores available in a machine. The threading structure is illustrated in figure 4. A natural division of work is employed, which lets one thread handle the GUI and graphical display, one thread the camera/projector calls, while a third is responsible for frame sequence decoding and a fourth for point cloud reconstruction. Practically, work is performed

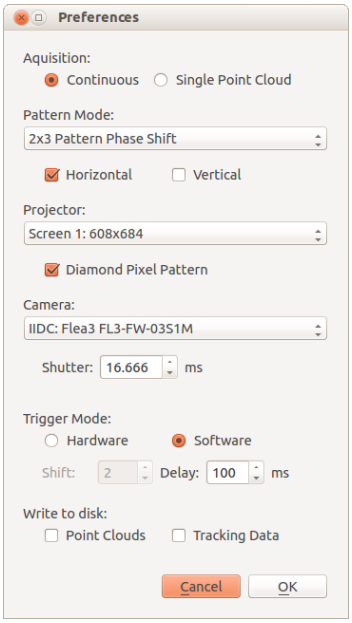


Fig. 3. Preference window showing pattern mode selection and camera/projector configuration.

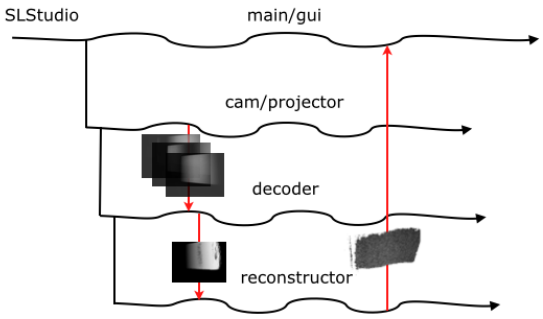


Fig. 4. Illustration of threads and inter-thread communication in SLStudio.

on heap-allocated arrays, and thread-communication consists of passing *smart pointers* to these to other threads for further processing.

IV. Conclusion

In conclusion, we have introduced a feature-rich software framework for real-time structured light. In our use case, it enables dense scene acquisition with over 300.000 points at 20 Hz, and allows other users to build custom low cost depth cameras with real-time processing ability. It also provides a platform for research in structured light by allowing users to implement new pattern coding strategies. The software is available open-source, and it is our hope that the community finds it useful, and participates in the continued development of it.

References

[1] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 160(2):128–160, 2011.

[2] Marvin Lindner, Ingo Schiller, Andreas Kolb, and Reinhard Koch. Time-of-Flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12):1318–1328, December 2010.

[3] Kai Liu, Yongchang Wang, Daniel L Lau, Qi Hao, and Laurence G Hassebrook. Dual-frequency pattern scheme for high-speed 3-D shape measurement. *Optics express*, 18(5):5229–44, March 2010.

[4] Oline Vinter Olesen, Rasmus R Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen. Motion tracking for medical imaging: a nonvisible structured light tracking approach. *IEEE transactions on medical imaging*, 31(1):79–87, January 2012.

[5] R.J. Valkenburg and a.M. McIvor. Accurate 3D measurement using a structured light system. *Image and Vision Computing*, 16(2):99–110, February 1998.

[6] Jakob Wilm, Oline Vinter Olesen, and Rasmus Larsen. Accurate and Simple Calibration of DLP Projector Systems. *Proc. SPIE 8979*, March 2014.

[7] R Yang and M Pollefeys. Multi-resolution real-time stereo on commodity graphics hardware. *Computer Vision and Pattern Recognition*, 1:I–211–I–217, 2003.

[8] Song Zhang. *High-resolution, Real-time 3-D Shape Measurement*. PhD thesis, Stony Brook University, 2005.

[9] Yong Zhao and Gabriel Taubin. Real-time stereo on GPGPU using progressive multiresolution adaptive windows. In *GPU Computing Gems Emerald Edition*, pages 473–495. 2011.

Appendix D

An MRI Compatible Surface Scanner

Conference	ISMRM Scientific Workshop, Tromsø 11. July – 14. July 2014
Presentation	Poster Presentation
Published in	DTU Orbit
doi	10.1118/1.596566

The poster presents MRI compatibility in our SL system achieved through shielding of the system components and optical relay of camera and projector images as detailed in Ch. 5. The system successfully acquires point clouds inside the scanner bore, and MRI data shows no degradation in quality due to the presence of the system.

An MRI Compatible Surface Scanner

Oline V. Olesen[123], Jakob Wilm[12], Andre Van der Kouwe[3], Rasmus R. Jensen[1], Rasmus Larsen[1], and Lawrence L. Wald[3]

[1] *DTU Compute, Technical University of Denmark, Denmark*

[2] *Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet, University of Copenhagen, Denmark*

[3] *Athinoula. A. Martinos Center for Biomedical Imaging, Dept. of Radiology, Massachusetts General Hospital, Massachusetts, USA*

Purpose: Deterioration of images due to motion is a continuing problem in medical imaging, especially fMRI, diffusion and dynamic PET where SNR is challenging and motion can be confounded with physiological change. Many motion correction techniques have been suggested but no single device or method solves the problem in routine clinical use. In this work, we propose a hardware design for an optical surface scanner for in-bore monitoring and markerless tracking of surfaces. It is the first remote structured light scanner that transmits projected patterns onto the subject and captures their images through optical fiber bundles. This technology, together with fast surface reconstruction algorithms and algorithms that separate rigid and non-rigid transformations, allows e.g. real-time motion correction feedback without the use of MR navigators or optical markers.

Methods: The presented device is designed to produce real time in-bore surface scans. To achieve a compact, RF noiseless, and low attenuation device, the electronics are separated from the optical end by two 670×500 image fibers of 2.7 m. Only a minimum of components are located in-bore, while the potentially RF emitting and ferromagnetic components are kept out of the bore.

The system consists of four parts: 1) a computer located outside the magnet room, 2) a power management unit, also outside the scan room, 3) a system box comprising a CCD camera with a near infrared (NIR) sensitive sensor, a DLP light projector, reengineered to project invisible NIR light, and optical data extenders, 4) two image fibers, and camera optics which extend from the system box into the magnet bore and transfer the projection image into the bore and the captured image out.

The system box is shielded for RF and is placed behind the magnet. It is custom made with a wooden frame covered by a 1 mm copper shield. The system box also contains fiber-optic data connections to and from the computer. 5/12 V DC power is supplied through capacitive filters, to avoid high frequency RF noise.

The system was set up on the Siemens mMR Biograph 3T scanner to demonstrate its feasibility in the narrow in-bore geometry further challenged by the limited view of the standard mMR head coil. The MRI compatibility of our system was initially tested by acquiring MPRAGE and EPI sequences of a phantom with the surface scanner: off/out-bore, on/out-bore, and on/in-bore, respectively.

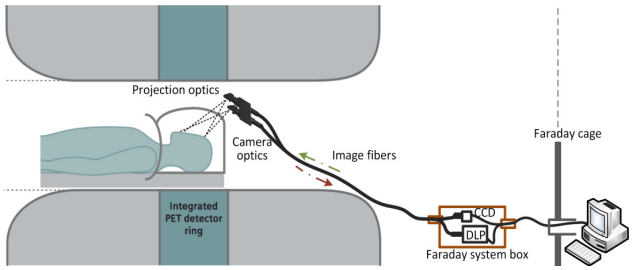


Fig. 1. Drawing of the developed structured light system integrated with the mMR Biograph. In-bore: the optical end facing a subject's head through the coil. Out-bore: optic fibers transmit images to and from the RF shielded system box (located outside if permitted by the room layout).

Results and Discussion: Fig. 2 shows two snapshot surface scans of a volunteer obtained using our device. The reconstructed point clouds are based on our non MRI compatible system[1]. Fig. 3 shows the effect on a MPRAGE sequence of having the surface scanner off/out-bore (left) compared to on/in-bore (mid). The difference image (right) shows no signs of suceptibility induced distortion. The effect on the uniformity is investigated by determining the SNR from two EPI volumes with equal sequence parameters[2]. The SNRs are 227.3 (off/out-bore), 227.7 (on/out-bore), 226.3 (on/in-bore), i.e. no impact on image SNR.

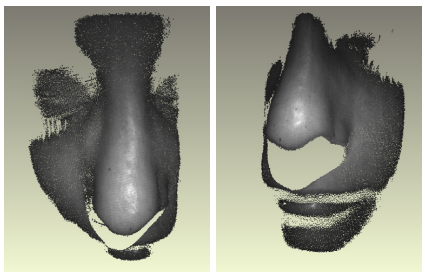


Fig. 2. In-bore surface scans of a volunteer's face (point clouds with texture overlay).

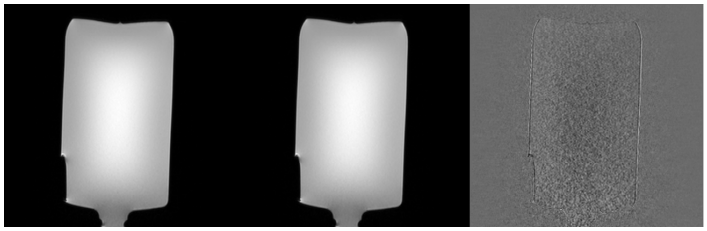


Fig. 3. MPRAGE images of a cylindrical phantom. Left to right: surface scanner off/out-bore, on/in-bore, difference of the first two images (normalized scale as the original scale

Conclusion: We have presented a remote surface scanner suitable for in-bore applications. The system design was shown to be MRI compatible and functional on the Siemens mMR Biograph.

References:

1. O. V. Olesen, J. M. Sullivan, T. Mulnix, R. R. Paulsen, L. Højgaard, B. Roed, R. E. Carson, E. D. Morris, and R. Larsen. "List-mode PET motion correction using markerless head tracking: proof-of-concept in human studies," *IEEE TMI,* vol. 32, 200-209 (2013).

2. Quality assurance methods and phantoms for magnetic resonance imaging: Report of AAPM nuclear magnetic resonance Task Group No. 1, *Medical Physics,* 17, 287 (1990); doi: 10.1118/1.596566.

Appendix E

Correction of Motion Artifacts for Real-Time Structured Light

Conference	Scandinavian Conference on Image Analysis, Copenhagen 15. – 17. July 2015
Presentation	Oral Presentation
Published in	Volume 9127 of the series Lecture Notes in Computer Science page 142-151
doi	10.1007/978-3-319-19665-7

In this contribution, an approach is presented for the correction of motion in real time SL. This problem is discussed in Sec. 3.6. The motion is estimated by means of fast phase correlation based image registration between sequences and corrected to reduce artifacts significantly, as demonstrated on a recording of a moving mannequin head. The approach is one of few current solutions to the motion problem.

Correction of Motion Artifacts for Real-Time Structured Light

Jakob Wilm[1,2], Oline V. Olesen[1,2], Rasmus R. Paulsen[1] and Rasmus Larsen[1]

[1] Dept. of Applied Mathematics and Computer Science,
Technical University of Denmark
Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby, Denmark
http://compute.dtu.dk/
[2] Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet,
Copenhagen University Hospital, University of Copenhagen

Abstract. While the problem of motion is often mentioned in conjunction with structured light imaging, few solutions have thus far been proposed. A method is demonstrated to correct for object or camera motion during structured light 3D scene acquisition. The method is based on the combination of a suitable pattern strategy with fast phase correlation image registration. The effectiveness of this approach is demonstrated on motion corrupted data of a real-time structured light system, and it is shown that it improves the quality of surface reconstructions visually and quantitively.

Keywords: 3D vision, structured light, motion correction, registration

1 Introduction

Structured light techniques are very popular for 3D scene capture, and a large variety of variants have been proposed. See [5] for an overview. The general idea is to aid in the detection of stereo correspondences by active projection of light onto the scene, followed by triangulation, to yield a surface reconstruction.

In time-multiplexed or multi-pattern structured light [13], a sequence of patterns is projected to generate one surface reconstruction, which encodes scene points, thereby disambiguating the matching process. The flexible nature of of structured light allows to choose a trade-off between number of patterns and accuracy, and to optimise for the radiometric properties of the scene.

With advances in light projection technology and computing power, it is today possible to perform structured light surface scanning with multiple patterns in real time. This enables accurate object tracking, dynamic deformation studies, fast object digitisation and many other applications. A general limitation with any multi-pattern method however is the underlying assumption of no motion between the patterns of a single sequence. A violation of this assumption often leads to large artifacts in the reconstructed surface. Such artifacts are shown as an example in figure 1.

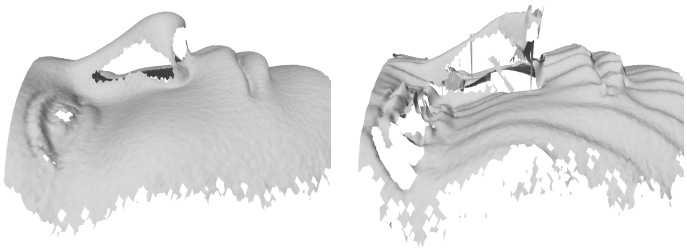


Fig. 1. Reconstructions of a phantom head using our structured light setup, showing the effects of motion during time-multiplexing. Left: static scene. Right: slight movement during acquisition.

By abandoning time multiplexing, single-shot structured light can be realised [14][6]. In the case of Microsoft Kinect 1, a static pseudorandom pattern is used. With single-shot techniques, the motion problem is avoided, but the reconstruction quality is also not very high or lateral resolution is lowered. Multiplexing by means of wavelength is also possible, i.e. by placing different patterns in the red, green and blue channels of the projection. This is not as robust to surface reflectance properties and spectral bleeding/crosstalk may occur. A review of single-shoot methods is provided in [20].

Time multiplexing remains the most robust technique, and is used in many commercial products, mainly for industrial inspection and reverse engineering. The chosen pattern strategy is important for the performance of the system, and two classes of methods remain very popular: binary Gray-coding and phase shifting profilometry (PSP). The former tends to be robust, but generally requires relatively many patterns, worsening the effects of motion. PSP can encode the scene unambiguously with only three patterns of a single sinusoid phase, shifted by 0 deg, 120 deg and 240 deg respectively. Depth resolution can be improved by using more shifts, or multiple phases, which gives a limited ambiguity that can be resolved using a phase-unwrapping algorithm.

In the presented method, we use the modified "2+1" phase shifting method of Zhang and Yau [19], which according to the authors reduces the effects of motion, and perform fast image registration on the acquired camera frames, to correct for the synchronisation error, and vastly improve the quality of scene reconstructions. By employing phase-based image registration, the motion correction is efficient enough to run in real-time in our 20 Hz surface scanning pipeline.

2 Previous work

Employing more than three patterns in phase shifting profilometry allows for sanity checking, and masking corrupted output [9]. This is at the expense of additional patterns in the sequence, generally making it slower (but more accurate), but it is questionable to what degree the unmasked output can be trusted.

In [7], a motion corrected real-time method is presented, for one-dimensional known motion.

The consequences of motion can be alleviated to some extend by running at high frequencies. Zhang et al. have presented structured light with dithered binary sinusoid patterns running at 1000 Hz reconstructions per second [15]. Their method requires strong projector lighting and high-speed cameras due to the short integration time of any single camera frame, and involves quality trade-offs making it unsuitable for many applications.

Liu et al. [10] propose motion corrected light with binary patterns and estimation of a global velocity vector based on the reconstructions, while Lu et al. show the theoretical feasibility of motion correction by means of image alignment [11].

3 Structured light and PSP

We use direct codification structured light, i.e. using a single camera-projector pair as described in [16]. Projector pixel coordinates are denoted (u_p, v_p), while camera pixel coordinates are (u_c, v_c). Employing a vertical baseline, the standard 3-step PSP algorithm encodes the projected images as

$$I_n^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\left(\frac{u_p}{N_p} - \frac{n}{3}\right)\right) \quad ,$$

where n indicates the pattern index, $n \in 1 \dots 3$ and N_p the number of projector columns. The camera captures the n'th pattern as

$$I_n^c(u_c, v_c) = A^c + B^c\cos\left(2\pi\left(\frac{u_p}{N_p} - \frac{n}{3}\right)\right) \quad .$$

Note the dependency on the projectors horizontal coordinate, u_p. A^c is the scene intensity including ambient contributions, while B^c is the modulation by projector light. A^c can be considered the magnitude of the Fourier DC component, while B^c is the magnitude at the principle frequency, and $\frac{u_p}{N_p}$ its normalised phase. In order to create correspondences, $\frac{u_p}{N_p}$ is extracted with the Fourier transform and scaled by N_p.

Our camera and projector pair are calibrated according to the method described in [18]. Corresponding sets of u_p, u_c, v_c-coordinates are used to triangulate object surface points as shown in figure 2. We denote one collection of projected patterns a pattern sequence, and the corresponding camera frames a frame set.

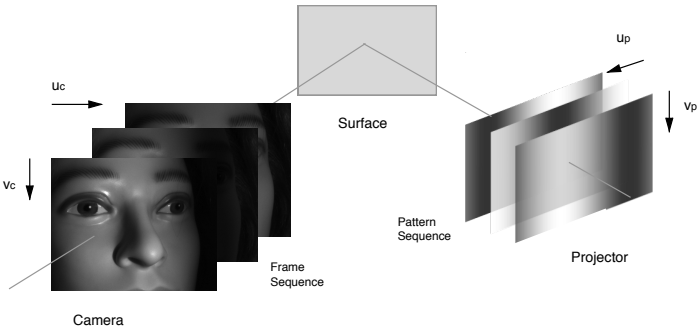


Fig. 2. Principle of structured light 3d scanning shown with the 3-step PSP pattern strategy. For each image point, the corresponding projector column coordinate, u_p, is extracted, which allows for triangulation of points on the object surface.

The "2+1" phase shifting method is a slight modification of the 3-step method, in which one of the patterns is a constant lit image:

$$I_1^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{u_p}{N_p}\right) \tag{1}$$

$$I_2^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{u_p}{N_p} - \frac{\pi}{2}\right) \tag{2}$$

$$I_3^p(u_p, v_p) = \frac{1}{2} \tag{3}$$

From the corresponding camera frame set, the horizontal projector coordinate can be extracted from camera frames as

$$u_p = \frac{N_p}{2\pi}\,\mathrm{atan2}\left(I_2^c - I_3^c, I_1^c - I_3^c\right) \quad ,$$

while intensity information is extracted simply as $A^c = I_3^c$.

The availability of these direct intensity frames, which are not corrupted by inter-frame motion, allows us to estimate the motion, and perform a correction. Specifically, we register intensity frames between frame sets, to estimate in 2D a global motion trend, $\boldsymbol{s} = [\Delta x, \Delta y]$. All frames of a set are then corrected by interpolating from the motion trend. This removes most of motion induced artifacts from single object sources that undergo rigid motion.

4 Phase correlation

In order estimate inter-frame motion, intensity frames are registered to each other. While the image deformation is most accurately described by a dense warp

field, any estimation method for such would be computationally prohibitive in a real-time context. We therefore utilise phase correlation, also called the Fourier-Mellin method, which can estimate a rigid transformation very quickly. It is a long well known method that is very noise-robust and particularly efficient [8][3], lending itself to real-time processing such as video stabilisation [4]. We use it here to register images of two successive frame sets, in order to correct for the misalignment of frames and the resulting artifacts in the 3D reconstruction.

In many applications, such as pose tracking, the tracked object is considered to undergo rigid 6 DOF motion between frames. In such scenarios, fast phase correlation registration can account for most of the misalignment, and reduce artifacts considerably.

With two images, f_1, f_2, of dimensions $N_x \times N_y$ the phase correlation technique first applies a spatial discrete Fourier transform to both images. The determination of translational shift is based on the Fourier shift theorem.

Assuming only a circular translational shift, the images are given as

$$f_2(x, y) = f_1(\mathrm{mod}(x - \Delta x,\, N_x), \mathrm{mod}(y - \Delta y,\, N_y)) \quad ,$$

and their respective Fourier spectra are related by

$$\mathcal{F}_2(u, v) = \; \mathrm{e}^{-2\pi i(u\Delta x/N_x + v\Delta y/N_y)} \mathcal{F}_1(u, v) \quad ,$$

with (u, v) denoting frequency components.

The normalised cross-power spectrum is given by

$$P = \frac{\mathcal{F}_1 \odot \mathcal{F}_2^*}{\|\mathcal{F}_1 \odot \mathcal{F}_2^*\|} = \; \mathrm{e}^{2\pi i(u\Delta x/N_x + v\Delta y/N_y)} \quad ,$$

with $\mathcal{F}^*$ denoting the complex conjugate, and $\odot$ the elementwise/Hadamard product. The cross correlation of f_1 and f_2 is now calculated as the inverse Fourier transform of P, and the shift estimated as its peak position

$$\boldsymbol{s} = \arg\max_{x,y} \mathcal{F}^{-1}(P).$$

Since the discrete Fourier shift theorem holds only for circular shifts, a window function is used on the Fourier transforms. In fact, this is usually necessary, as the edges on tiled input images provide strong high-frequency features in Fourier domain.

To obtain the peak position with sub-pixel precision, the centroid of the cross-correlation peak is computed by also including values from a small neighbourhood around $\boldsymbol{s}$.

It is also possible to recover scale and rotation parameters between images in an analogue way by converting magnitude spectra to the log-polar domain before computing the cross-power spectrum [3].

5 Experimental setup

Our structured light setup consists of a single camera-projector pair. The camera integration time is a multiple of the projector refresh period for truthful gray-value reproduction, and a hardware trigger signal ensures accurate synchronisation of camera and projector. The projector update frequency is 120 Hz, and the camera integration time 8.333 ms. Due to trigger latency, we are able to capture every other projector image, resulting in $60\,\mathrm{s}^{-1}$ camera frames and 20 Hz surface reconstructions per second using the 2+1 phase shifting method.

With two consecutive frame sets $I_t^c = I_{1,t}^c, I_{2,t}^c, I_{3,t}^c$ and $I_{t+1}^c = I_{1,t+1}^c, I_{2,t+1}^c, I_{3,t+1}^c$ at times t and $t+1$ respectively, we perform phase correlation based registration between the flat intensity frames $I_{3,t+1}^c$ and $I_{3,t}^c$. This shift, $\boldsymbol{s}$, serves as the global motion estimate at that time point. Before reconstructing a surface from frame set I_{t+1}^c, its frames are corrected to:

$$I_{1,t}^c\left(\boldsymbol{x} - \frac{2}{3}\boldsymbol{s}\right) \quad I_{2,t}^c\left(\boldsymbol{x} - \frac{1}{3}\boldsymbol{s}\right) \quad I_{3,t}^c\left(\boldsymbol{x}\right)$$

As noted above, phase correlation in the log-polar domain allows for determination of global scale and rotation. However, in most cases of small-scale misalignment caused by camera or object movement, a translation registration is adequate and preferred by us due to the lower computational demands.

We carry out experiments by scanning a phantom head that is moving in a controlled manner. It is mounted on a stepper motor that rotates back and forth on a 36 deg arc at a constant speed of $11,4\,\mathrm{deg/s}$ at a distance of approximately 20 cm to the camera. In our camera, this results in per-coordinate pixel shifts of up to 5 px between to consecutive frames. This scenario is modelled after our head-tracking approach used for medical imaging motion correction [12].

To quantify the quality of our motion corrected structured light approach, we perform a tracking experiment on the phantom head. The head is held stationary to capture an uncorrupted reference scene. The motor is then started, and the object surface is scanned and registered to the reference by means of the iterative closest point algorithm. We use the point-to-plane error metric [2], as it generally converges best on this kind of data [17]. Correspondences are found by means of back-projecting data into the camera frame [1]. The so-obtained tracking data is used to evaluate the quality of the reconstructed object surface by means of the root-mean-square (RMS) error of the alignment to the reference surface. Partial overlap of surfaces is handled by removing those point correspondences that match to the border of the reference point cloud.

6 Results

The alignment of image frames I_1^c and I_3^c during the motor controlled motion scene is shown in figure 3. It is seen, that the global shift was estimated correctly (see e.g. edge of iris, corners of the eye), and a large amount of misalignment between the images is accounted for.

Fig. 3. Registration of camera frames I_1^c (green colour channel) and I_3^c (magenta colour channel) belonging to the same frame sequence. Left: before registration. Right: after registration. The misalignment is most easily seen at the edge of the iris and corners of the eye.

The effect of motion correction on the resulting surface reconstruction is shown in figure 4.

The correction shows reduction of artifacts and distortion, especially around the nose area.

Using the rotating phantom head described, we use phase correlation to estimate the translational shift between intensity frames over time. This is shown in figure 5 with the raw shifts, and their cumulative sums. From these plots, it becomes apparent that phase correlation faithfully and reproducibly captures the global object movement over time.

RMS errors for the moving scene are shown in figure 6.

The static scene trace is influenced solely by image noise in the camera frames. With corrected motion, the error increases by approximately 7% (time-averaged RMS), while in the uncorrected case, the increase in error is over 19%. It is noted that the increase in RMS is mainly due to large artifacts in high-frequency regions of the camera frame. It is seen that our motion correction approach effectively decreases these artifacts, and lowers the RMS alignment error.

The average processing time for the alignment of two camera intensity frames ($300 \times 200\,\mathrm{px}$) was around 5 ms.

7 Conclusion and discussion

We have proposed a fast alignment strategy to reduce motion artifacts in time-multiplexed structured light. While the method is limited to recovering a global translational shift, this was shown to explain a large amount of the misalignment, and hence reduce artifacts in the resulting surface reconstructions. The used method was used for its robustness and computational efficiency. With its fixed-time performance, the phase correlation technique is suitable for real-time processing, as done in our implementation.

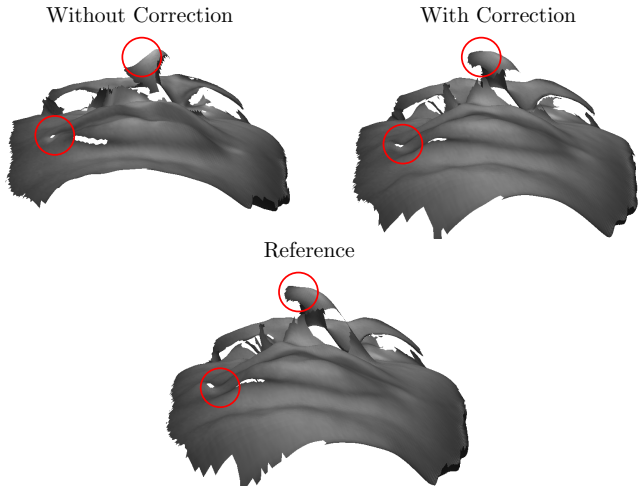


Fig. 4. Results of performing phase correlation based motion correction on an object moving at constant speed. Two aliased regions are highlighted in red circles. Top left: without correction. Top right: with correction. Bottom: stationary reference scan, e.g. no motion.

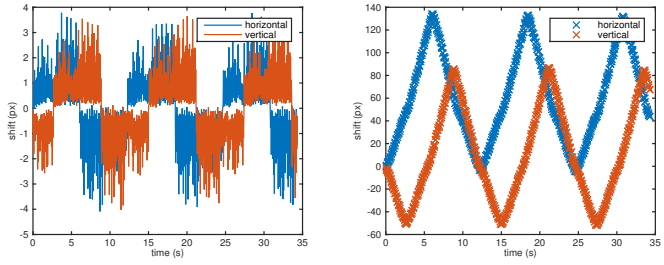


Fig. 5. Left: Translational shift estimates in horizontal (blue) and vertical (red) directions. Right: The cumulative sum of shift estimates.

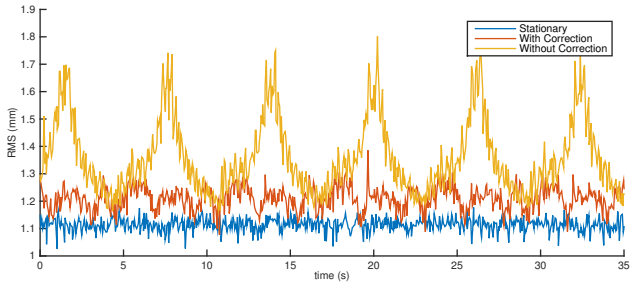


Fig. 6. RMS error of surface alignment to stationary scene at $t = 0$ using ICP. Variation in the stationary case is due to image noise. It is seen that motion corrected frame sets result in markedly lower RMS. In the uncorrected case, RMS peaks when the object is farthest from stationary situation. Mean RMS over time: stationary (1.12 mm), with correction (1.20 mm), without correction (1.34 mm).

The limitation of this method is that it only accurately explains single object translational movement. With offline-processing, a large variety of video stabilisation techniques can be combined with the 2+1 phase shifting method. These include, but are not limited to subframe correlation methods, feature-based parametric image registration and optical flow based methods.

For many situations in which structured light is used for tracking or object digitisation, our method promises to significantly improve reconstruction results.

References

1. G. Blais and M.D. Levine. Registering multiview range data to create 3D computer objects. *IEEE PAMI*, 17(8):820–824, 1995.
2. Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. *Image and vision computing*, 1992.
3. E. De Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE PAMI*, 9(5):700–703, 1987.
4. S. Erturk. Digital image stabilization with sub-image phase correlation based global motion estimation. *IEEE Transactions on Consumer Electronics*, 49, 2003.
5. J. Geng. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 160(2):128–160, 2011.
6. C. Guan, L. Hassebrook, and D. Lau. Composite structured light pattern for three-dimensional video. *Optics express*, 11(5):406–17, March 2003.
7. I. Ishii, T. Koike, and T. Takaki. Fast 3D shape measurement using structured light projection for a one-directionally moving object. *IECON 2011*, pages 135–140, 2011.
8. C.D. Kuglin and D.C. Hines. The phase correlation image alignment method. *IEEE International Conference on Cybernetics and Society*, volume , pages 163–165, 1975.

9. D. Lau, K. Liu, and L.G. Hassebrook. Real-time three-dimensional shape measurement of moving objects without edge errors by time-synchronized structured illumination. *Optics letters*, 35(14):2487–9, 2010.
10. Y. Liu, H. Gao, Q. Gu, T. Aoyama, T. Takaki, and I. Ishii. A Fast 3-D Shape Measurement Method for Moving Object. *Int. Conf. Progress in Informatics and Computing (PIC)*, pages 219-223, 2014.
11. L. Lu, J. Xi, Y. Yu, and Q. Guo. New approach to improve the accuracy of 3-D shape measurement of moving object using phase shifting profilometry. *Optics Express*, 21(25):30610–30622, 2013.
12. O.V. Olesen, R.R. Paulsen, L. Hjgaard, B. Roed, and R. Larsen. Motion tracking for medical imaging: a nonvisible structured light tracking approach. *IEEE Transactions on Medical Imaging*, 31(1):79–87, 2012.
13. J.L Posdamer and M.D Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17, 1982.
14. K. Sakashita, Y. Yagi, R. Sagawa, R. Furukawaa, and H. Kawasaki. A system for capturing textured 3D shapes based on one-shot grid pattern with multi-band camera and infrared projector. *Proc. 3DIMPVT*, pages 49–56, 2011.
15. Y. Wang, J.I. Laughner, I.R. Efimov, and S. Zhang. 3D absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique. *Optics express*, 21(5):6631–6636, 2013.
16. J. Wilm, O.V. Olesen, and R. Larsen. SLStudio : Open-Source Framework for Real-Time Structured Light. *Proc IPTA 2014*, pages 1–4. IEEE Xplore, 2014.
17. J. Wilm, O.V. Olesen, R.R. Paulsen, L. Hjgaard, B. Roed, and R. Larsen. Real time surface registration for PET motion tracking. *LNCS: SCIA 2011*, 6688 LNCS:166–175, 2011.
18. J. Wilm, O.V. Olesen, and R. Larsen. Accurate and Simple Calibration of DLP Projector Systems. *SPIE Photonics West 2014*, pp. 897909, 2014.
19. S. Zhang and S.T. Yau. High-speed three-dimensional shape measurement system using a modified two-plus-one phase-shifting algorithm. *Optical Engineering*, 46(November 2007):1–6, November 2007.
20. Z.H. Zhang. Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques. *Optics and Lasers in Engineering*, 50(8):1097–1106, 2012.

Appendix F

Method for Surface Scanning in Medical Imaging and Related Apparatus

---

Patent No.	WO 2015/071369 A1
Priority Date	Nov 13, 2013
Publication Date	May 21, 2015

This patent relates to the use of optical fiber bundles for the purpose of MRI compatible SL and motion correction. It covers our implementation of an SL system with details regarding the optical coupling of image fiber bundles to a projector source and a camera, as described in Ch. 5.1.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 May 2015 (21.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/071369 A1

(51) International Patent Classification:
A61B 5/00 (2006.01) *A61B 5/055* (2006.01)
A61B 5/11 (2006.01) *G02B 6/32* (2006.01)
A61B 6/00 (2006.01)

(21) International Application Number:
PCT/EP2014/074509

(22) International Filing Date:
13 November 2014 (13.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13192786.5 13 November 2013 (13.11.2013) EP

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 1, DK-2800 Kgs. Lyngby (DK).

(72) Inventors: **OLESEN, Oline**; Gladsaxevej 78, st.tv., DK-2860 Søborg (DK). **LARSEN, Rasmus**; Ermelundsvej 123, DK-2820 Gentofte (DK). **WILM, Jakob**; Borgmestervangen 4A, DK-2200 Copenhagen N (DK). **JENSEN, Rasmus Ramsbøl**; L.I. Brandes Allé 4, 2th, DK-1956 Frederiksberg C (DK).

(74) Agent: **ZACCO DENMARK A/S**; Arne Jacobsens Allé 15, DK-2300 Copenhagen S (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) Title: METHOD FOR SURFACE SCANNING IN MEDICAL IMAGING AND RELATED APPARATUS

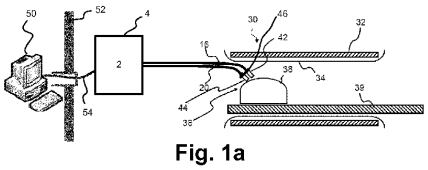


Fig. 1a

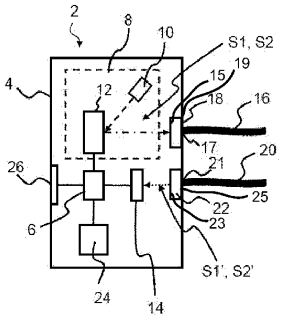


Fig. 1b

(57) Abstract: A method and apparatus for surface scanning in medical imaging is provided. The surface scanning apparatus comprises an image source, a first optical fiber bundle comprising first optical fibers having proximal ends and distal ends, and a first optical coupler for coupling an image from the image source into the proximal ends of the first optical fibers, wherein the first optical coupler comprises a plurality of lens elements including a first lens element and a second lens element, each of the plurality of lens elements comprising a primary surface facing a distal end of the first optical coupler, and a secondary surface facing a proximal end of the first optical coupler.

WO 2015/071369 A1

METHOD FOR SURFACE SCANNING IN MEDICAL IMAGING AND RELATED APPARATUS

The present invention relates to a method and apparatus for surface scanning in medical imaging, in particular in magnetic resonance imaging (MRI), in positron emission tomography (PET), and/or in combined MRI/PET. The invention may be used for surface scanning/motion tracking in particular inside small geometries (in-bore of PET, MRI, CT, SPECT or combined scanners as PET/CT and MRI/PET).

BACKGROUND

Over the last decade, numerous methods for surface scanning and motion tracking in brain imaging have been developed, but head motion during scanning pertains to be a significant problem causing artefacts and significantly reducing image quality.

Known methods include external tracking systems as well as image based motion tracking and correction. Many external tracking systems use markers attached to the subjects head. This potentially introduces errors and complicates the process of preparing the subject for the scan and therefore reduces the usability in clinical practice. Correspondingly, the image based motion tracking methods developed for medical brain imaging generally suffer from an inability to obtain sufficiently high temporal and spatial resolution at the same time. Further, the high resolution of modern medical scanners (down to tenths of a millimeter for MRI and a few millimeters for PET) set strict requirements to motion tracking systems.

SUMMARY

The present invention relates to a method and apparatus for improved surface scanning in medical imaging. Disclosed herein is therefore a method for surface scanning in medical imaging that may be used for subject tracking, the method comprising a) providing an image source and a first fiber bundle comprising first optical fibers having proximal ends and distal ends; b) positioning the distal ends of the first optical fibers within a scanner borehole of a medical scanner; c) feeding an image from the image source into a proximal end of a first optical coupler, the first optical coupler comprising a plurality of lens elements including a first lens element and a second lens element; and d) feeding an image from a distal end of the first optical coupler into the proximal ends of the first optical fibers.

Disclosed herein is also a surface scanning apparatus for surface scanning in medical imaging, the apparatus comprising a) an image source, b) a first optical fiber bundle comprising first optical fibers having proximal ends and distal ends, and c) a first optical

coupler for coupling an image from the image source into the proximal ends of the first optical fibers, wherein the first optical coupler comprises a plurality of lens elements including a first lens element and a second lens element, each of the plurality of lens elements comprising a primary surface facing a distal end of the first optical coupler, and a secondary surface facing a proximal end of the first optical coupler.

By the above method and/or surface scanning apparatus is obtained an improved surface scanning method and/or motion tracking method wherein components that generate noise, such as radio emitting components and/or ferromagnetic components, are separated form and kept out of the bore. Further, occlusion effects are highly reduced if not completely avoided. Further, an improved image quality on the object which is scanned in the borehole is provided. Problems previously observed regarding a decrease in image quality due to long distances between scanner and light source is avoided due to the use of optical fibers, which ensures a high image quality even over larger distances.

The method may be particularly useful in a method for motion tracking in medical imaging, and the surface scanning apparatus may be a motion tracking apparatus

By the method and/or surface scanning apparatus is further obtained a very compact device, which can easily be incorporated into a scanner or be used as an add-on to existing scanning systems.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other features and advantages of the present invention will become readily apparent to those skilled in the art by the following detailed description of exemplary embodiments thereof with reference to the attached drawings, in which:

Fig. 1a schematically illustrates a surface scanning apparatus in connection with a medical scanner and a computer system,

Fig. 1b schematically illustrates an exemplary surface scanning apparatus,

Fig. 2 schematically illustrates parts of an exemplary surface scanning apparatus,

Fig. 3 schematically illustrates parts of an exemplary surface scanning apparatus,

Fig. 4 schematically illustrates parts of an exemplary surface scanning apparatus,

Fig. 5 schematically illustrates parts of an exemplary surface scanning apparatus,

Fig. 6a schematically illustrates decreasing of the image size with different lens elements in an optical coupler,

Fig. 6b schematically illustrates increasing of the image size with different lens elements in an optical coupler,

Fig. 7a schematically illustrates a relay lens coupler, and

Fig. 7b schematically illustrates an alternative relay lens coupler.

DETAILED DESCRIPTION

The figures are schematic and simplified for clarity, and they merely show details which are essential to the understanding of the invention, while other details may have been left out. Throughout, the same reference numerals are used for identical or corresponding parts.

Surface scanning incorporates tracking spatial position of a surface or surface points over time and/or tracking/determining spatial position of a surface or surface points at a given time.

The medical scanner may be a magnetic resonance (MR) scanner. Further, the method and apparatus for motion tracking may be employed for motion correction of scanning images obtained by other medical scanners, such as a positron emission tomography (PET) scanner, a single photon emission computed tomography (SPECT) scanner or a computed tomography (CT) scanner. In one or more aspects, the method and apparatus may be employed for motion correction of a subject in a combined PET-MR scanner or a combined PET-CT scanner.

The image source provided in the method or the apparatus may include a light source and/or a digital micromirror device (DMD) chip, where the DMD chip is for modulating the incoming light from the light source thus creating a pre-determined image source.

The image source may be a modified DLP (digital light processing) projector.

Feeding an image, e.g. from the image source into a proximal end of a first optical coupler and/or from a distal end of the first optical coupler into the proximal ends of the first optical fibers, may comprise feeding a pattern sequence comprising a pattern or a plurality of different patterns.

The image source may be configured for providing a pattern sequence, e.g. comprising a plurality of different patterns, e.g. for projection of patterns onto the surface region or scene of the subject in the borehole. A pattern sequence (S), e.g. a first pattern sequence (S1) and/or a second pattern sequence (S2), comprises one or more patterns (P), such as a plurality of different patterns including a primary pattern and a secondary pattern. A pattern sequence comprises or consists of a number N of

patterns. A pattern sequence may be defined by pattern sequence parameters, for example including number of patterns, configuration/structure of respective patterns, order of patterns and/or timing of pattern(s) of the pattern sequence. The duration of a pattern sequence may be in the range from 1 millisecond to about 1 second. The duration of a pattern sequence may be about 10 milliseconds, about 20 milliseconds, about 50 milliseconds, about 100 milliseconds or about 200 milliseconds.

A pattern may comprise a number of pixels, e.g. arranged in an array along a first and second axis. A pattern may be defined by pattern parameters, e.g. including pixel settings (color/wavelength and/or intensity) of each pixel and/or one or more groups of pixels in the pattern. A group of pixels of a pattern may be referred to as a subregion denoted R of a pattern. Accordingly, a pattern may comprise one or more subregions R_1, R_2, R_3..., a subregion comprising one or more pixels. Pattern sequence parameters may include pattern parameters, e.g. of a primary pattern, a secondary pattern and/or a tertiary pattern.

The image source may comprise a light modulator.

The light modulator or DMD chip can be adapted for projection of patterns onto the surface region or scene of the subject in the borehole. The light modulator may comprise a liquid crystal display (LCD) chip or a DMD chip. In one or more embodiments, the light modulator may comprise a liquid crystal on silicon (LCOS) chip. In one or more embodiments, the light modulator may comprise grids, slits or filters. The light modulator may be a transmitting or reflective light modulator.

The DMD chip/light modulator may be an array which is approximately 9.86 mm times 6.16 mm and images from the DMD chip/light modulator are mapped with the first optical coupler into a first fiber bundle with proximal end size of about 6.7mm times 5mm.

The image source may be connected to a control unit for receiving control signal(s) from the control unit. The control signal(s) may comprise pattern sequence parameters, such as number, configuration, order and/or timing of pattern(s) of the pattern sequence. In one or more embodiments, the control signal(s) may comprise a pattern sequence selector, and the image source may be configured for projecting different pattern sequences dependent on the pattern sequence selector.

The resolution of the image source and/or first fiber bundle limits the pattern resolution projected onto the subject. The image source may have a resolution of at least 500 pixels, such as at least 1,000 pixels or at least 10,000 pixels in order to project a useful

image on the subject. In an exemplary method and/or apparatus, the image source may have a resolution of HVGA (480x320 pixels) or more, e.g. (608x684 pixels), SVGA(800x600 pixels), XGA (1024x768 pixels), 720p (1280x720 pixels), or 1080p (1920x1080 pixels).

In one or more embodiments, a number of different pattern sequences may be stored in the image source, and the image source may be configured to project a selected pattern sequence based on a pattern sequence selector from a control unit.

In an embodiment, the light source may include one or more lasers or (high power) LED's including a first laser/LED configured to emit light at the first wavelength λ_1 and/or a second laser/LED configured to emit light at a second wavelength λ_2. The light source may also include a third laser/LED configured to emit light at a third wavelength λ_3.

The light source may include a broad spectrum light source, such as a metal-halide lamp. In one or more embodiments, the light source may comprise a light emitting diode (LED). The light source may comprise a filter for forming light with desired frequency spectrum/wavelength distribution. In one or more embodiments, the light source may be adapted to emit light in the infrared (IR) or near-infrared (NIR) range, for example at a wavelength in the range from 700 nm to about 1,000 nm, e.g. about 850 nm. In one or more embodiments, the light source may be adapted to emit light in the UV range.

In one or more embodiments, the image source may comprise light at a first wavelength λ_1 in the range from 780-900 nm. For example, the wavelength range may be between 800-860 nm. The first laser/LED may be a red or orange/red laser, wherein the first wavelength λ_1 is in the range from about 590 nm to about 700 nm. In one or more embodiments the first wavelength λ_1 is about 635 nm. The first laser/LED may be an LED, wherein the first wavelength λ_1 is in the range from about 830 nm to about 870 nm, e.g. about 850 nm or from about 810 nm to about 850 nm. The first laser/LED may be an LED, wherein the first wavelength λ_1 is in the range from about 790 nm to about 830 nm, e.g. about 810 nm or from about 800 nm to about 820 nm.

The second laser/LED may be a green laser, wherein the second wavelength λ_2 is in the range from about 490 nm to about 560 nm, e.g. about 532 nm. The second laser/LED may be an LED, wherein the second wavelength λ_2 is in the range from about 880 nm to about 920, e.g. about 900 nm.

The third laser/LED may be a blue or violet laser, e.g. wherein the third wavelength λ_3 is in the range from 430 nm to about 490 nm, e.g. about 445 nm or about 473 nm. The third laser/LED may be an LED, e.g. wherein the third wavelength λ_3 is in the range from 930 nm to about 1,000 nm, e.g. about 940 nm.

The light source may comprise a UV source, e.g. configured to emit light with a wavelength in the range from about 230 nm to about 400 nm, e.g. about 350 nm.

One or more mirrors or a prism may be used to guide light or an image from the light source and/or image source to the first optical coupler. Different examples of this are shown and described in connection with fig. 2-5.

The first optical coupler may comprise or consist of an even number of lens elements, e.g. two, four, six, eight, ten, twelve or more lens elements. In one or more embodiments, ten lenses are included in the first optical coupler. In another embodiment, six lenses are included in the first optical coupler. When choosing a lower number of lenses, the optical loss is kept at a minimum, whereas when choosing a many lenses, the image quality is improved and the distortion and blurriness are reduced. The relay lens element may comprise between four and twelve lens elements.

The first optical coupler may be adapted for either increasing or decreasing the size of the image after the image has passed through the first optical coupler. In an exemplary method/apparatus, the lens elements in the first optical coupler maps the incoming image size by a ratio of 1:1.2, thus the image size of the image coming out of the distal end of the first optical coupler is 20% larger compared to the size of the image entering the first optical coupler at its proximal end. In general, the image size can be mapped in the range from 1:0.5 (i.e. the out-coming image is 50 % smaller than the incoming image) to 1:2.

Advantageously, the first optical coupler may be a relay lens coupler.

The distal end of the first optical coupler may be secured releasably to the proximal end of the first fiber bundle by a click-release-coupling. This allows for an easy and flexible positioning of the optical fibers in the borehole of the scanner or an easy replacement and/or exchange of the optical fibers or the first optical coupler without moving the other of the two.

Alternatively, for ensuring a constant optimum coupling of the image from the first optical coupler into the optical fibers, the distal end of the first optical coupler may be fixed non-releasably to the proximal end of the first fiber bundle.

A second optical coupler comprising a plurality of lens elements including a first lens element and a second lens element may also be included in the surface scanning apparatus and/or the method for tracking the motion. Also, a second fiber bundle comprising second optical fibers having proximal ends and distal ends can be provided and its distal ends positioned within the scanner borehole of the medical scanner. The distal ends of the second optical fibers may be applied for capturing a projected image from a subject in the borehole. This projected image will normally be fed from the proximal ends of the second optical fibers into the second optical coupler.

The second optical coupler may also be adapted for either increasing or decreasing the size of the projected image after the image has passed through the second optical coupler.

At least one of the plurality of lens elements in the first and/or second optical coupler may be achromatic.

In an embodiment of the invention, the first lens element in the first and/or second optical coupler can be positioned at the proximal end of the first and/or second optical coupler, respectively, and the second lens element can be positioned at the distal end of the first and/or second optical coupler, respectively. The first lens element and the second element may further be achromatic with convex sides pointing towards each other.

The primary surface of each of the plurality of lens elements in the first and/or second optical coupler may be concave or convex or planar or a combination thereof. Likewise, the secondary surface of each of the plurality of lens elements in the first and/or second optical coupler may be concave or convex or planar or a combination thereof. The primary surface of one or more lens elements may be concave. The primary surface of one or more lens elements may be convex. The primary surface of one or more lens elements may be plane. The secondary surface of one or more lens elements may be concave. The secondary surface of one or more lens elements may be convex. The secondary surface of one or more lens elements may be plane.

The apparatus and the method may further comprise a mirror and/or a prism, and light from the light source may pass the mirror/prism before entering the first optical coupler. The first optical fibers may further be adapted for projecting at least one pattern from the image source via the first optical fibers onto the surface region of the subject positioned in a borehole of the medical scanner.

The first optical fibers may comprise at least 100 optical fibers, such as at least 10,000 fibers, each fiber corresponding to a pixel in a pattern projected onto the surface region of the subject. In one or more embodiments, the number of first optical fibers is equal to or larger than the number of pixels in the image source, for full benefit of the image source resolution. The number of first optical fibers may match or be in the range of ±20% of the resolution of the image source. In one or more embodiments, the number of first optical fibers is less than the number of pixels in the image source, for full benefit of the optical fibers.

The second optical fibers can be adapted for capturing at least one projected pattern and/or image projected form the subject. The second optical fibers may comprise at least 100 optical fibers, such as at least 100,000 fibers. Each second optical fiber may correspond to one or more pixels in a first camera, which the captured image is transmitted to. In one or more embodiments, the number of second optical fibers is equal to or larger than the number of pixels in the first camera for increasing the processing time of the camera. In one or more embodiments, the number of second optical fibers is less than the number of pixels in the first camera for increasing the precision of the image capturing. The number of second optical fibers may match or be in the range of ±50% of the resolution of the first camera.

The first camera may be a CCD camera or a CMOS camera. The first camera may have a resolution of at least 640x480, e.g. 1280X960, 3264x2448 or more.

The surface region may have an area of at least 0.1 cm^2, e.g. in the range from 1 cm^2 to 500 cm^2. In one or more embodiments, the surface region area may be in the range from 20 cm^2 to 100 cm^2.

The surface region may at least partly cover a nasal region of the subject. This may lead to improved motion tracking due to the significant curvature of the subject surface in this region. Further, facial movements are limited near the bridge of the nose which is preferred when tracking the motion of the scull and the brain.

The apparatus may also comprise a first lens assembly, i.e. projector side projection optics, arranged at and/or attached to the distal end of the first optical fibers for coupling images or pattern sequences from the first optical fibers to the surface region of the subject. The distal ends of the second optical fibers may be provided with a second lens assembly, i.e. image capturing optics, for coupling images or pattern sequences from the surface region of the subject to the second optical fibers.

The apparatus may comprise a frame, wherein the first and second lens assemblies are mounted on the frame. The frame fixes the position between the two distal ends of the fibers bundles and/or between the first and second lens assemblies to maintain a fixed positional relationship in order to provide an accurate movement correction and/or such that the two fiber bundles can be moved together inside the borehole. The distal ends of the first and the second fiber bundles may be mounted on the frame.

The first and second lens assemblies will normally comprise an objective lens with a given focal length and an aperture. The focal length may be changed by exchanging the objective lens. Also by changing the distance between the objective lens and the distal ends of the optical fibers in the first or second optical fiber bundle, control of how much of the image source illuminates the subject and which area projected light is captured from, respectively, can be obtained. The aperture in the first and/or second lens assemblies may be adjusted by opening and/or closing them, which also provides a tool for controlling the output from the first optical fibers onto the subject, and the projected image from the subject into the second optical fibers for the first lens assembly and the second lens assembly, respectively.

The second lens assembly may also comprise a filter, e.g. a NIR filter. Likewise, the first lens assembly may also comprise a filter.

The first lens assembly may comprise a first mirror/prism. The second lens assembly may comprise a second mirror/prism, respectively. A common mirror/prism may be shared between the first lens assembly and the second lens assembly. A mirror/prism in a lens assembly may provide redirection of the light which may lead to larger freedom in positioning the distal fiber ends/lens assemblies in the bore.

The first and second optical fibers may be arranged in respective first and second fiber arrays. In one or more embodiments, the first optical fibers may comprise a first array of at least 10,000 fibers, such as 100x100 fibers, such as 400x400 or 600x600 fibers or 680x480 fibers or 1,200x1,200 fibers or more. The first optical fibers may comprise a first array of at least 100,000 fibers, e.g. 5,000x5,000 fibers. In one or more embodiments, the second optical fibers comprise a second array of at least 10,000 fibers, such as 100x100 fibers, such as at least 400x400 or 600x600 fibers or 680x480 fibers or 1,200 x1,200 fibers, or more. The second optical fibers may comprise a second array of at least 100,000 fibers, e.g. 5,000x5,000 fibers. The optical fibers may be arranged in an array of any suitable size and shape, e.g. rectangular, circular, oval, polygonal or others. Typically, the fiber diameter is in the range from 5 to 20 micrometers. The number of first optical fibers may be larger than 1,000, such as larger

than 10,000. The number of second optical fibers may be larger than 1,000, such as larger than 10,000.

Using first and second optical fibers enables or facilitates the use of the method and apparatus for medical scanners with a permanent magnetic field surrounding the object, e.g. an MR scanner. Further, using first and second optical fibers enables or facilitates the use of the method and apparatus for medical scanners with limited access to the subject due to the subject being positioned in a scanner borehole during scanning.

The first and second fiber bundles may each have a length larger than 1 meter, such as larger than 2 meters, e.g. about 5 meters or about 10 meters. In an exemplary apparatus and/or method, the first and second fiber bundles may each have a length between 1 and 5 meters, such as between 2.5 and 3 meters, for example about 2.7 meters. Having a length of the fiber bundles in this length range may enable the user to place the distal ends of the first and second fiber bundles inside the scanner while keeping the first and/or second optical couplers at a different location remote from the scanner or even remote from/outside the scanning room.

The length of the fiber bundles allows for positioning of a power management part and/or a computer for controlling a sequence, an image pattern or similar relating to the image source, outside the room with the scanner. This allows for the creation of a remote surface scanner. By separating the electronics from the optical end by the two fibers bundles, a compact, radio frequency noiseless and low attenuation surface scanner is achieved.

In the apparatus and method of this invention, a minimum of components are located in the borehole of the scanner and the disturbing components are kept outside the borehole. This maintains the field of view and the high resolution of nowadays surface scanners. Further, the components located in the borehole of the scanner may be made of non-metallic materials.

The surface scanning apparatus may also comprise a housing which surrounds all the motion tracking elements apart from the fibers bundles which extend from the surface of the housing. The housing may be a radio frequency shielded box costume normally made out of a frame covered by a thin copper layer or sheet, e.g. of a thickness of 1 mm. Any metal suited for shielding the electric components may be used.

A filter of one or more capacitors may ensure that the electromagnetic noise from powering the components inside the housing does not propagate along the power

cable. Correspondingly, a power supply is positioned outside the scanner room and the power is led though a filter in the wall into the scanner room and the inside of the housing to feed the relevant components therein.

The surface scanning apparatus may be constructed such that it is part of the medical scanner or be used as an add-on to existing scanners.

Fig. 1a schematically illustrates a medical scanner 30 for use with the method and apparatus. The scanner 30 is an MR scanner comprising a permanent magnet 32 in a scanner housing 34 forming a scanner borehole 36. The scanner 30 comprises a head coil 38 for scanning a subject positioned on the support structure (scanner bed) 39.

First lens assembly 42 and second lens assembly 44 are mounted to respective distal ends of first optical fibers 16 and second optical fibers 20 and positioned in the scanner borehole 36 for projecting and detecting pattern sequences on/from a surface region within the head coil 38.

As an alternative to the MR scanner shown in fig. 1a, PET scanner comprising at least one detector ring in a scanner housing forming a scanner borehole could also be imagined. In this case, the distal ends of the respective optical fibers 16, 20 could be positioned outside the detector ring and near the scanner borehole for projecting and detecting pattern sequences on/from a surface region within the scanner borehole. Yet an alternative to the MR scanner of Fig. 1a is a combined MR/PET scanner.

Fig. 1a shows a surface scanning apparatus 2 which is positioned inside the scanner room defined by surrounding walls 52 illustrated by one wall/Faraday cage to the left side of the apparatus 2. A power management and/or controller part 50, e.g. a computer as illustrated in Fig. 1a, is positioned outside the scanner room. The surface scanning apparatus 2 may be positioned outside the scanner room defined by surrounding walls 52 if the optical fibers 16, 20 are sufficiently long.

In Fig. 1a is also shown an optical extender 54 which transfers image data noiseless between the surface scanning apparatus 2 and the computer 50 outside the scanner room. The apparatus 2 can be surrounded by a housing 4 which functions as a radio frequency shielded box. The housing 4 can be made out of a frame, e.g. a wooden frame, covered by a 1 mm copper layer. A filter of capacitors (not shown in the figure) ensures that the electromagnetic noise from powering the components inside the housing does not propagate along the power cable. The power supply optionally being a separate power supply or a part of the power management/controller part 50 is

positioned outside the scanner room and the power is led through a filter in the wall 52 into the scanner room and the elements inside the housing 4 of the apparatus 2.

The distal ends of fibers are provided with respective first and second lens assemblies 42, 44 constituting projection optics and image capturing optics, respectively. A frame 46 is used for fixing the position between the first and second lens assemblies 42, 44 and/or between the distal ends of the first and second optical fibers 16, 18, respectively.

The first and second lens assemblies may each comprise an objective lens with a given focal length and aperture. Also, the second lens assembly may comprise a near infra-red (NIR) filter. Both first and second lens assemblies may comprise a first mirror/prism and/or second mirror/prism, respectively. The mirror/prism may be shared between the two lens assemblies.

Fig. 1b schematically shows a surface scanning apparatus 2 of the present invention. The apparatus 2 comprises a housing 4 accommodating a control unit 6 and an image source 8 comprising a light source 10 and a light modulator 12. Further, the apparatus 2 optionally comprises a first camera 14 connected to the control unit 6 for exchange of control signals and/or pattern sequence data between the control unit 6 and the first camera 14. During use, first optical fibers 16 are coupled to the apparatus at the proximal ends 17 of the first optical fibers via first optical coupler 18 such that light from the image source 8 is coupled into the first optical fibers 16. The first optical coupler 18 has a proximal end 15 and a distal end 19.

The apparatus optionally comprises a memory unit 24 and a user interface unit 26.

The first optical fibers 16 may be fixedly mounted to the housing 4, i.e. the first optical fibers 16 may form a part of the apparatus 2. Alternatively, a distal end 19 of the first optical coupler 18 may be secured releasably to the proximal ends 17 of the first fiber bundle 16 by a click-release-coupling.

During use, second optical fibers 20 are coupled to the apparatus 2 at the proximal ends 21 of the second optical fibers 20 via second optical coupler 22 such that pattern sequences or images projected on the surface region is detected by the first camera 14. The second optical coupler 18 comprises a proximal end 23 and a distal end 25.

The first and second optical fibers may be fixedly mounted to the housing 4, i.e. the first and second optical fibers may form a part of the apparatus 2, thereby simplifying setting up the apparatus.

Alternatively, the distal end 19 of the first optical coupler 18 and/or the distal end 25 of the second optical coupler 22 may be secured releasably to the proximal ends 17 of the first fiber bundle 16 and the proximal ends 21 of the second fiber bundle 20, respectively, by a click-release-coupling.

The apparatus 2 is configured for projecting a first pattern sequence (S1) onto a surface region of the subject with the image source 10, wherein the subject is positioned in a scanner borehole of a medical scanner, the first pattern sequence optionally comprising a first primary pattern ($P_{1,1}$) and a first secondary pattern ($P_{1,2}$). The apparatus 2 may be configured for detecting the projected first pattern sequence (S1') with the first camera 14. The control unit 6 optionally determines a second pattern sequence (S2) comprising a second primary pattern ($P_{2,1}$) based on the detected first pattern sequence (S1') and sends control signals to the image source 8 with image source 10 and light modulator 12 projecting images in the form of the second pattern sequence (S2) onto a surface of the subject via the first optical coupler 18. The projected second pattern sequence (S2') may be detected with the first camera 14 and the pattern sequence data are processed in the control unit and/or in the first camera 14 and/or in external computer 50. Upon or during detection of pattern sequence data, the apparatus 2 or external computer 50 determines motion tracking parameters based on the detected second pattern sequence (S2').

Figs. 2-5 show different embodiments of the first optical coupler 18 comprising a plurality of lens elements $l_{1,...,}\ l_N$, including a first lens element l_1 and a second lens element l_2. In Fig. 2, two lens elements are provided whereas Figs. 3 and 4 show a large plurality of lens elements. In Fig. 5, the first optical coupler 18 is a relay lens coupler comprising or consisting of a number of N lens elements positioned inside an outer housing of the relay lens coupler. N may be six, eight or ten.

Figs. 2-5 show only the first optical coupler 18, however the second optical coupler 22 may have an identical or different construction as the embodiments shown in Figs. 2-5 for the first optical coupler 18. The following description of the lens elements in the first optical coupler 18 may therefore also apply to the lens elements in the second optical coupler 22.

Each lens element of the plurality of lens elements $l_{1,...,}\ l_N$ comprises a primary surface 28 facing a distal end 19 of the first optical coupler 18, and a secondary surface 29 facing a proximal end 15 of the first optical coupler 18. Normally, there will be an even number of lens elements in the first and/or second optical coupler 18, 22. There may be two, four, six, eight, ten, twelve or more lens elements $l_{1,...,}\ l_N$.

One or more of the lens elements l_N may be achromatic, e.g. at least one of the plurality of lens elements is achromatic. In Fig. 2-5 only chromatic lens elements are shown.

In one or more embodiments, the first lens element l_1 is positioned at the proximal end 15 of the first optical coupler 18 and the second lens l_2 element is positioned at the distal end 19 of the first optical coupler 18, as shown in Fig. 2. In Fig. 2, both lens elements are chromatic. However, the first lens element l_1 and the second element l_2 could also be achromatic with convex sides pointing towards each other.

In the apparatus, mirrors and/or prisms may be used to guide the image from the image source 8 to the first optical coupler 18. In Figs. 2 and 3, a mirror 7 is used for guiding the image from the image source to the light modulator 12 from where it is guided to the proximal end 15 of the first optical coupler 18. In Figs. 4 and 5, the image passes from the image source 8 through a prism 9 to the light modulator 12 from where it again passes through the prism 9 in such a manner that the image is guided directly into the proximal end 15 of the first optical coupler 18.

The first and/or second optical coupler 18, 22 may be adapted for either increasing or decreasing the size of the image and/or the projected image, respectively such that the size of the image / projected image is either larger or smaller after having passed through the first and/or second optical coupler.

A simple schematic illustration of how the image size can be increased or decreased using an optical coupler is shown in Fig. 6a-b. In Fig. 6a, the image size is decreased from a size d_{in} of the incoming image to a size of d_{out} of the outcoming image, where $d_{in} > d_{out}$, whereas in Fig. 6b, the image size is increased from d_{in} of the incoming image to a size of d_{out}, where $d_{in} > d_{out}$. The different focal length f_1, f_2 of the lens elements are illustrated in the figures.

By utilizing more than two lens elements, an improved correction and reduced (geometric) distortion may be obtained. Further, aberration effects are reduced. This allows the user to control the how large a part of the image which is coupled into the first optical fibers 16 and control the size of the projected image, which comes out of the second optical coupler 22 after having been collected by the second optical fibers 20.

Figs. 7a-b show two different examples of a relay lens couplers which may be used in the invention as the first optical coupler 18 and/or the second optical coupler 22.

In Fig. 7a, the relay coupler comprises or consist of six lens elements l_1, l_2, l_3, l_4, l_5, l_6 arranged symmetrically such that the outermost lens elements l_1, l_6 are nearly identical in size oriented such that they are a mirror image each other. Likewise, the lens elements l_2, l_5 positioned adjacent to the outermost lens elements l_1, l_6 form a mirror image pair and so forth for the next lens elements approaching the middle of the relay lens coupler. Four of the lens elements l_1, l_3, l_4, l_6 are planoconvex, i.e. they have a convex side and a plane side, whereas the other two lens elements l_2, l_5 are biconcave, i.e. both the primary and the secondary side of the lens elements are concave.

Fig. 7b shows a relay lens coupler comprising ten lens elements l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8, l_9, l_{10} again arranged symmetrically with the lens elements pair wise from the two outermost lens elements towards the centre of the relay lens coupler being mirror images of one another. In Fig. 7a, four of the lens elements l_1, l_4, l_7, l_{10} are planoconvex, two of the lens elements l_3, l_8 are biconcave, two of the lens elements l_2, l_9 are biconvex, i.e. both the primary and the secondary side of the lens elements are convex, and the last two elements l_5, l_6 are planoconcave, i.e. they have a concave side and a plane side.

The number of lens elements pairs is not limited to the examples shown in Fig. 7a-b. Further the combination of sizes and shapes of the lens elements may also vary, e.g. different combinations of planoconcave, planoconvex, biconcave, and/or biconvex lens element pairs positioned such they form a mirror image of one another could also be imagined.

It should be noted that in addition to the exemplary embodiments of the invention shown in the accompanying drawings, the invention may be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the concept of the invention to those skilled in the art.

REFERENCES

2	Apparatus
4	Housing
6	Control unit
7	Mirror
8	Image source
9	Prism
10	Light source
12	Light modulator
14	First camera
15	Proximal end of the first optical coupler
16	First optical fibers
17	Proximal ends of first optical fibers
18	First optical coupler
19	Distal end of the first optical coupler
20	Second optical fibers
21	Proximal ends of second optical fibers
22	Second optical coupler
23	Proximal end of the second optical coupler
24	Memory
25	Distal end of the second optical coupler
26	User interface
28	Primary surface of the lens elements
29	Secondary surface of the lens elements
30	Medical scanner
32	Magnet
34	Scanner housing
36	Scanner borehole
38	Head coil
39	Scanner bed
40	Subject
42	First lens assembly
44	Second lens assembly
46	Frame
50	Power management part
52	Wall surrounding the scanner room

54	Optical extender
l_N	N'th lens element
l_1	First lens element
l_2	Second lens element
l_3	Third lens element
l_4	Fourth lens element
l_5	Fifth lens element
l_6	Sixth lens element
l_7	Seventh lens element
l_8	Eights lens element
l_9	Ninths lens element
l_{10}	Tenths lens element
d_{in}	Size of the image before entering the first/second optical coupler
d_{out}	Size of the image after exiting the first/second optical coupler
f_1	Focal length of the first lens element
f_2	Focal length of the second lens element

CLAIMS

1. A method for surface scanning in medical imaging, the method comprising

providing an image source and a first fiber bundle comprising first optical fibers having proximal ends and distal ends;

positioning the distal ends of the first optical fibers within a scanner borehole of a medical scanner;

feeding an image from the image source into a proximal end of a first optical coupler, the first optical coupler comprising a plurality of lens elements including a first lens element and a second lens element;

feeding an image from a distal end of the first optical coupler into the proximal ends of the first optical fibers.

2. Method according to claim 1, wherein the image source has a resolution of 480x320 pixels or more.

3. Method according to any of claims 1-2, wherein the image source includes a digital micromirror device (DMD) chip.

4. Method according to any of claims 1-3, wherein the image source is connected to a control unit for receiving control signal(s) from the control unit, the control signal(s) comprising a pattern sequence selector, and wherein the image source is configured for projecting different pattern sequences dependent on the pattern sequence selector.

5. Method according to claim 4, wherein a number of different pattern sequences are stored in the image source, and the image source is configured to project a selected pattern sequence based on the pattern sequence selector from the control unit.

6. Method according to any of claims 1-5, wherein the image source comprises a light source, and a mirror or a prism is used to guide light from the light source towards the first optical coupler.

7. Method according to any of the preceding claims, wherein the first optical coupler is a relay lens coupler.

8. Method according to any of the preceding claims, wherein the distal end of the first optical coupler is releasably secured to the proximal end of the first fiber bundle by a click-release-coupling.

9. Method according to any of claims 1-7, wherein the distal end of the first optical coupler is fixed non-releasably to the proximal end of the first fiber bundle.

10. Method according to any of the preceding claims, the method comprising

providing a second optical coupler comprising a plurality of lens elements including a first lens element and a second lens element;

providing a second fiber bundle comprising second optical fibers having proximal ends and distal ends;

positioning the distal ends of the second optical fibers within the scanner borehole of the medical scanner;

capturing a projected image from a subject in the borehole by the distal ends of the second optical fibers; and

feeding the projected image from the proximal end of the second optical fibers into the second optical coupler.

11. Surface scanning apparatus for surface scanning in medical imaging, the apparatus comprising

an image source,

a first optical fiber bundle comprising first optical fibers having proximal ends and distal ends, and

a first optical coupler for coupling an image from the image source into the proximal ends of the first optical fibers,

wherein the first optical coupler comprises a plurality of lens elements including a first lens element and a second lens element, each of the plurality of lens elements comprising a primary surface facing a distal end of the first optical coupler, and a secondary surface facing a proximal end of the first optical coupler.

12. Surface scanning apparatus according to claim 11, wherein the image source has a resolution of 480x320 pixels or more.

13. Surface scanning apparatus according to any of claims 11-12, wherein the image source includes a digital micromirror device (DMD) chip.

14. Surface scanning apparatus according to any of claims 11-13, wherein the image source is connected to a control unit for receiving control signal(s) from the control unit, the control signal(s) comprising a pattern sequence selector, and wherein the image source is configured for projecting different pattern sequences dependent on the pattern sequence selector.

15. Surface scanning apparatus according to claim 14, wherein a number of different pattern sequences are stored in the image source, and the image source is configured to project a selected pattern sequence based on the pattern sequence selector from the control unit.

16. Surface scanning apparatus according to any of claims 11-15, wherein the first optical coupler comprises an even number of lens elements, e.g. two, four, six, eight, ten, twelve or more lens elements.

17. Surface scanning apparatus according to anyone of the claims 11-16, wherein the first optical coupler is a relay lens coupler.

18. Surface scanning apparatus according to anyone of the claims 11-17, wherein the distal end of the first optical coupler is releasably secured to the proximal end of the first fiber bundle by a click-release-coupling.

19. Surface scanning apparatus according to anyone of the claims 11-17, wherein the distal end of the first optical coupler is fixed non-releasably to the proximal end of the first fiber bundle.

20. Surface scanning apparatus according to anyone of the claims 11-19, wherein at least one of the plurality of lens elements is achromatic.

21. Surface scanning apparatus according to claim 20 wherein the first lens element is positioned at the proximal end of the first optical coupler and the second lens element is positioned at the distal end of the first optical coupler, wherein the first lens element and the second element are achromatic with convex sides pointing towards each other.

22. Surface scanning apparatus according to anyone of the claims 11-21, wherein the primary surface of each of the plurality of lens elements is concave or convex, and the secondary surface of each of the plurality of lens elements is concave or convex.

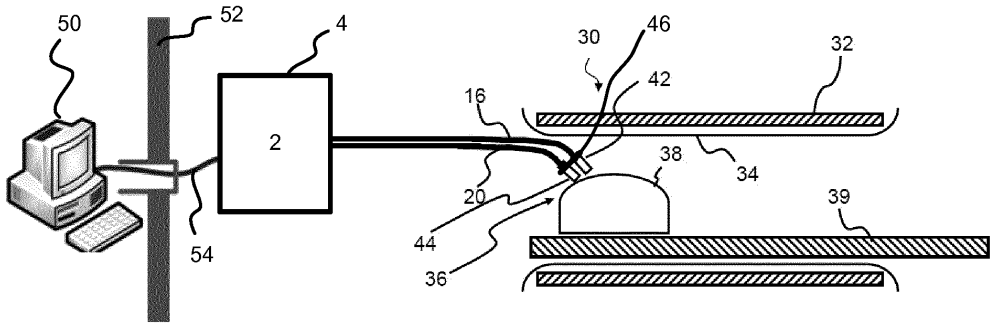

50
52
4
2
54
30
46
42
16
32
34
38
20
44
36
39

Fig. 1a

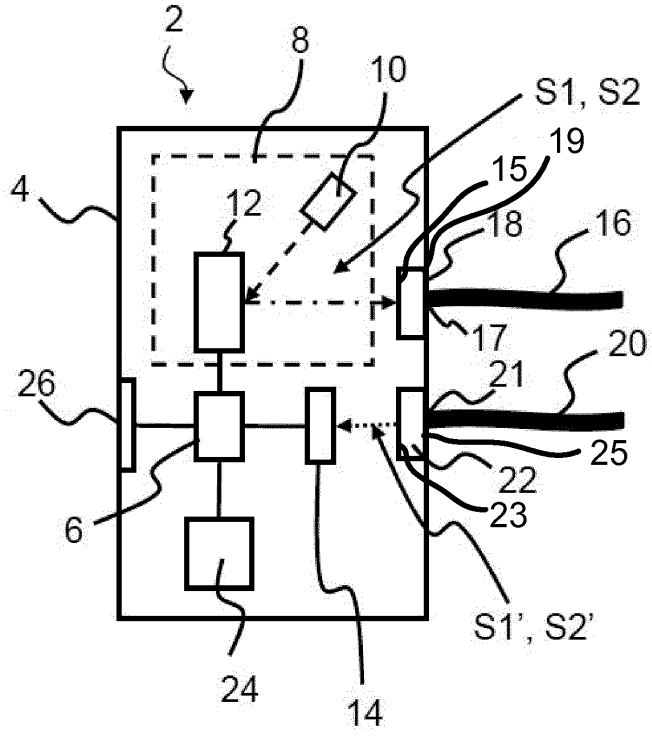

2
8
10
S1, S2
19
15
4
12
18
16
17
20
21
26
25
22
23
6
S1', S2'
24
14

Fig. 1b

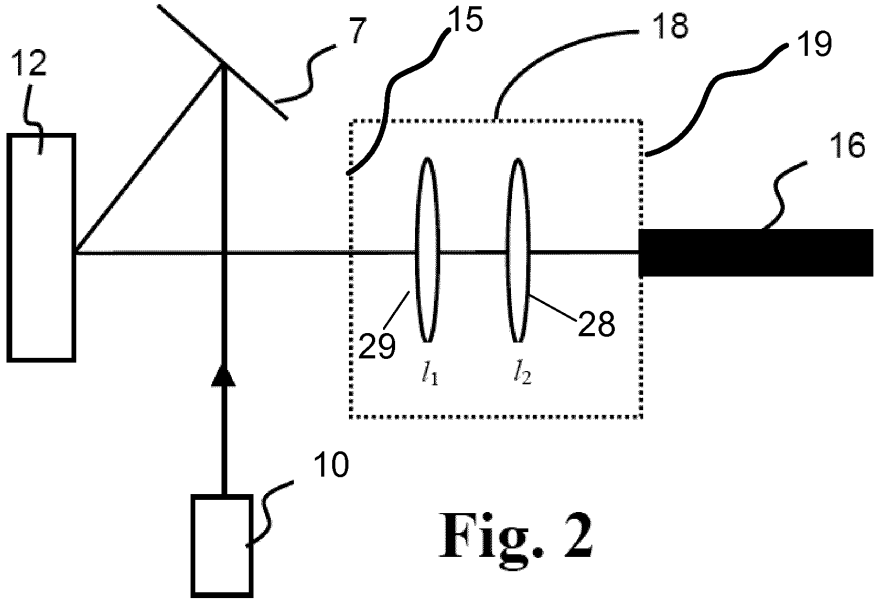


Fig. 2

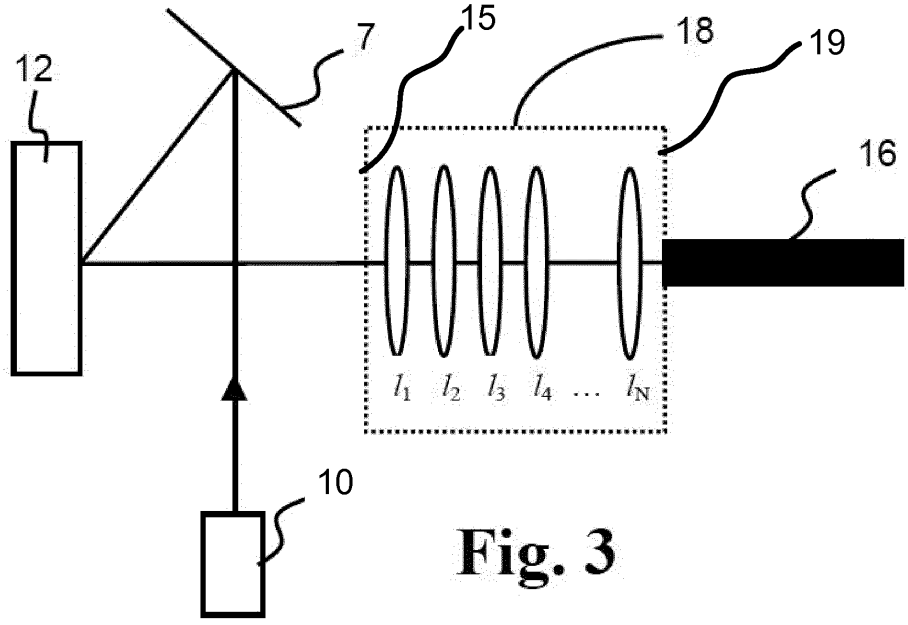


Fig. 3

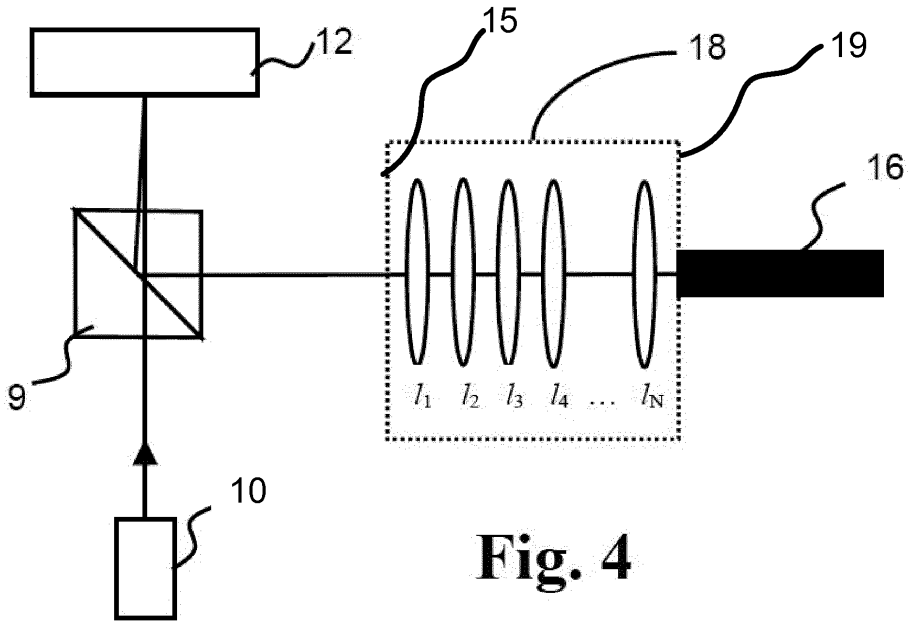


Fig. 4

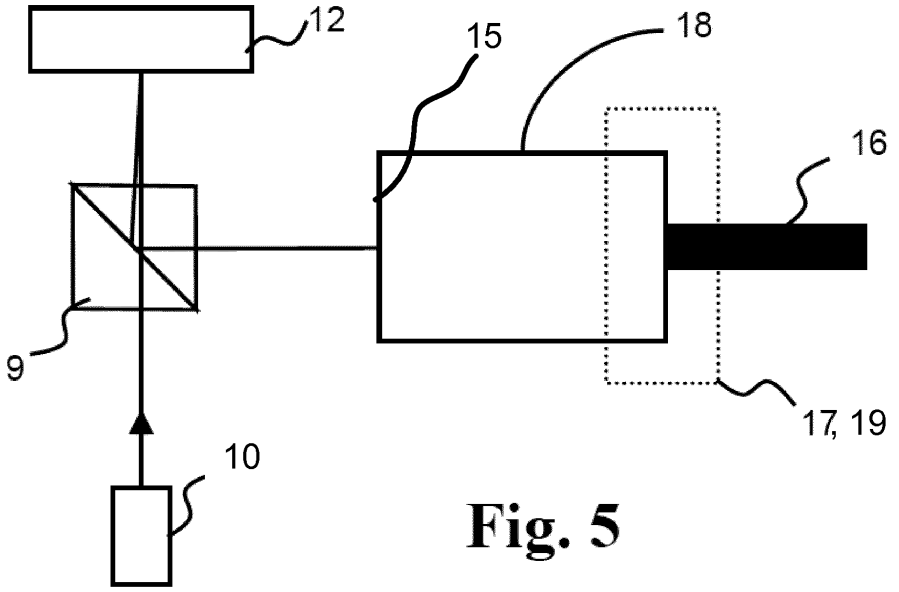


Fig. 5

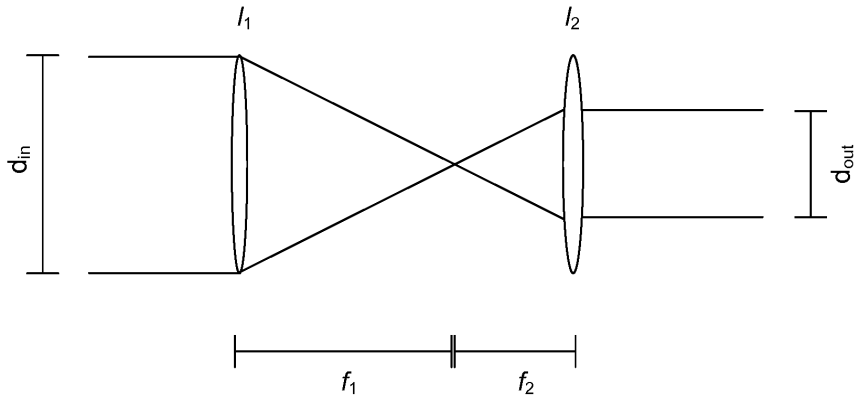


Fig. 6a

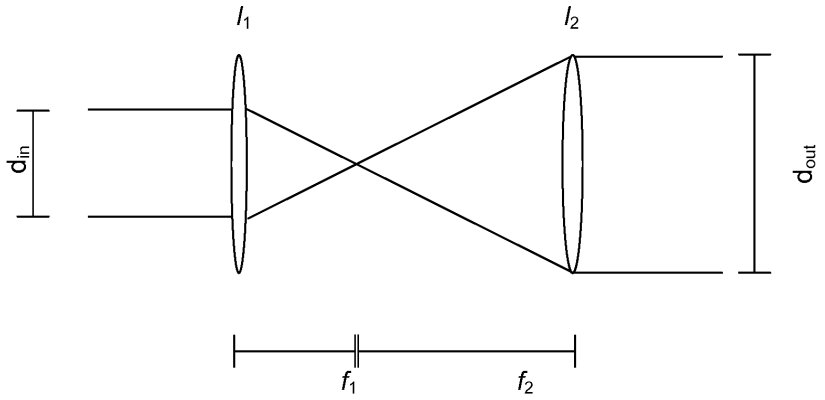


Fig. 6b

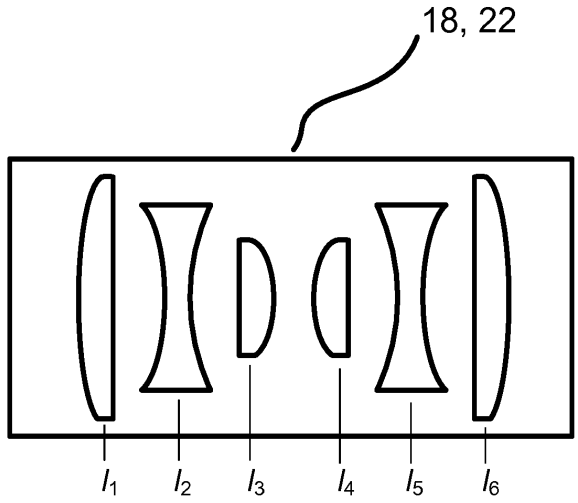


Fig. 7a

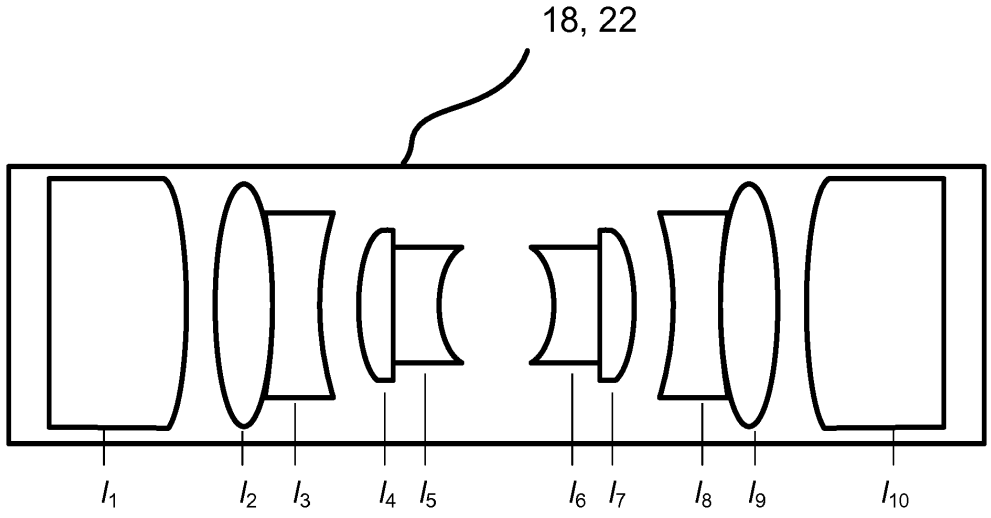


Fig. 7b

Appendix G

Precision and Accuracy Parameters in Structured Light 3-D Scanning

Conference	LowCost3D Sensors, Algorithms, Applications, Berlin 1. and 2. December 2015
Presentation	Oral Presentation
Published in	ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
doi	10.5194/isprsarchives-XL-5-W8-7-2016

In this contribution, we investigate the performance of a high-precision SL system with respect to calibration and scene coding parameters. The contribution lies in the systematic treatment and established evaluation using an industry standard. The parameters influencing accuracy are discussed in Ch. 6. While these results are obtained in a highly accurate system with reconstruction times in the order of seconds, they are valid also for real time systems.

PRECISION AND ACCURACY PARAMETERS IN STRUCTURED LIGHT 3-D SCANNING

Eyþór R. Eiríksson[a], Jakob Wilm[a], David B. Pedersen[b] and Henrik Aanæs[a]

[a] Department of Applied Mathematics and Computer Science
[b] Department of Mechanical Engineering
Technical University of Denmark
{eruei,jakw,aanes}@dtu.dk, dbpe@mek.dtu.dk

KEY WORDS: Structured Light, 3d Scanning, Accuracy Assessment, VDI 2634 (2)

ABSTRACT:

Structured light systems are popular in part because they can be constructed from off-the-shelf low cost components. In this paper we quantitatively show how common design parameters affect precision and accuracy in such systems, supplying a much needed guide for practitioners. Our quantitative measure is the established VDI/VDE 2634 (Part 2) guideline using precision made calibration artifacts. Experiments are performed on our own structured light setup, consisting of two cameras and a projector. We place our focus on the influence of calibration design parameters, the calibration procedure and encoding strategy and present our findings. Finally, we compare our setup to a state of the art metrology grade commercial scanner. Our results show that comparable, and in some cases better, results can be obtained using the parameter settings determined in this study.

1. INTRODUCTION

Structured Light (SL) systems enable robust high quality capture of 3D geometry, and are actively used throughout several fields. These systems can be constructed using commercial off the shelf (COTS) hardware, making them accessible and affordable. The obtainable accuracy and precision of such systems vary considerably, and are mainly functions of several design parameters. The influence of these parameters has not been studied extensively in the literature. Previously, no combined study has systematically investigated the effect of common parameter choices on the final result and quantified them using an established standard.

To address the lack of work in this regard, we investigate how common design choices influence precision and accuracy. Our analysis is based on our own active stereo-vision setup consisting of two industrial cameras and a consumer projector. We empirically show our parameter selection such that maximum performance is obtained, and quantify using the VDI/VDI 2634 (Part 2) guideline. Finally, we compare our results to a commercial metrology grade scanner (GOM ATOS III Triple Scan) as a benchmark against state of the art, with decent results. Throughout this study we seek to employ widely available and accepted methods & models used in such systems to obtain easily reproducible results.

The contribution of this paper lies in the attempt to quantitatively answer the following questions

- What calibration parameters should be included in the calibration procedure?
- What angular range of observations is required in the calibration procedure?
- How many observations are required for calibration?
- Which SL encoding strategy is the overall best performer?

We believe this to be valuable information for practitioners wanting to build their own system, *e.g.* as part of research projects or industrial implementations.

This paper is structured as follows. Section 2 covers related work. Section 3 gives an overview of our experimental setup. Section 4, 5 and 6 covers our investigations on calibration parameters, calibration observations and encoding strategies respectively. In section 7 we compare our system to a commercial system and finally, we conclude in section 8.

2. RELATED WORK

Much work has been devoted to the field of SL systems *e.g.* (Rocchini et al., 2001), (Fisher et al., 1999), (Chen et al., 2008), (Gühring, 2000a), (Tsai and Hung, 2005). These contributions have mostly dealt with the methodological development of such systems whereas less focus has been placed on quantitative accuracy and precision analysis. One of the most important factors with respect to accuracy is system calibration. While recent focus has been placed on projector-camera calibration (Zhang and Huang, 2006), (Li et al., 2008), (Legarda-Sa et al., 2004), we here consider an active stereo vision setup (Heikkila and Silvén, 1997), (Tsai, 1987, Zhang, 2000), without projector calibration. Precision is considered to be mostly dependent on the encoding strategy. A vast selection of methods have been proposed, see (Salvi et al., 2004), (Gorthi and Rastogi, 2010), (Geng, 2011) for recent surveys. While many of these methods aim to reduce the number of patterns, the amount of outliers and computational complexity, less focus has been placed on precision. Here, we compare selected encoding strategies from a precision and accuracy perspective.

Characterising SL systems in terms of accuracy is a challenging and ongoing problem, which despite its relevance has only seen few published guidelines and standards. The only currently published standard is the German VDI/VDE 2634 Part 2 guideline (2634, 2000), (Guidi, 2013), *Optical 3-D measuring systems – Optical systems based on area scanning*. This guideline aims to capture the complex nature of such a system, using a number of length and shape measurements throughout the scanning volume. Researchers have already accepted this guideline for evaluation of 3D scanning systems (Boehm, 2014), (Luhmann, 2011), (Beraldin and Gaiani, 2005), (Beraldin et al., 2007). We here argue that the guideline is lacking to some extent. Firstly, it fails to cap-

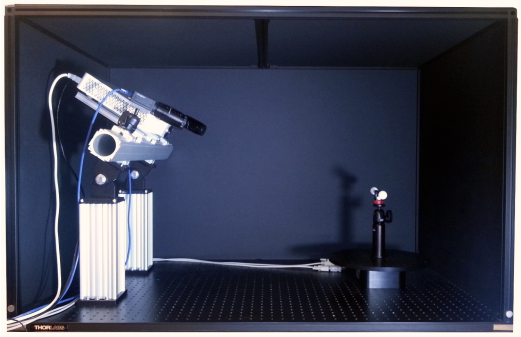


Figure 1: Our structured light system setup with two high-resolution industrial cameras, a Full HD LED projector and a rotation stage mounted on a rigid aluminum mount. Specifications are given in Table 1.

ture frequency response characteristics of SL systems using the proposed low frequency artifacts. Lastly, the artifacts are optically ideal for SL scanning. Therefore, results only indicate 'best case' results, given that particular material. The standard is however well suited for relative measures *e.g.* for acceptance testing and benchmarking purposes.

Limited work has been conducted on SL parameter investigations and their effect on overall performance (Lohry et al., 2009). However, to the authors knowledge, no quantitative evaluation has been performed on how the different SL parameters directly influence the final results as defined by the VDI/VDE guideline.

3. EXPERIMENTAL SETUP

Our structured light setup, as seen in Figure 1, consists of two industrial cameras (Point Grey Research GS3-U3-91S6C-C) and a high resolution DLP projector (LG PF80G) mounted on a rigid aluminum beam structure. Technical specifications are given in Table 1. In addition, a high precision turntable is used in order to provide automatic rotation of a calibration checkerboard. The apparatus and scan objects can be fully enclosed during capture, in order to prevent ambient light contamination.

Parameter	Specification
Cameras CCD	Sony ICX814 1"
Cameras Resolution	3376×2704 px
Camera Lens Focal Length	16 mm
Camera Lens Aperture	5.6
Camera Exposure	66.66 ms
Projector Resolution	1920×1080 px
Camera Baseline	450 mm
Camera Object Distance	~ 750 mm
Stereo Field of View (FOV)	$(300 \times 300 \times 230)$ mm

Table 1: Technical specifications of our structured light setup.

Figure 2 shows the calibration plate and Figure 3 shows the VDI/VDE 2634(2) measurement artifacts used during this study. The artifacts consist of a flat white painted aluminum plate and two ceramic spheres separated by a known distance. Both artifacts have been measured according to procedure T3-01 of ISM3D using a coordinate measurement machine (CMM), and traceability has been established through the virtual CMM method. Specifications for nominal values and attached uncertainties are listed in Table 2 and 3.

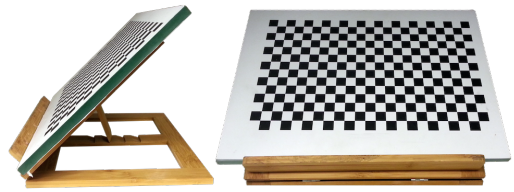


Figure 2: The calibration plate used in this study sitting on a rigid wooden support frame. Manufactured from $(400 \times 280 \times 12$ mm unhardened float-glass. A high resolution printed checkerboard is glued on the flat surface.

Figure 3: Calibration artifacts according to the VDI/VDE 2634(2) standard. Top: painted and lapped aluminum flat. Bottom: alumina-circonium ceramic spheres on a carbon-fiber rod. Nominal values are given in table 2 and 3.

Following the VDI/VDE 2634 (2), we use four quality parameters:

- Probing error form, P_F, which describes the radial range of residuals from a least squares fit sphere with up to 0.3% of the worst points rejected.
- Probing error shape, P_S, measuring the signed deviation between the least squares fit diameter and the nominal. Again, up to 0.3% of the worst points are rejected.
- Sphere distance error, SD, denoting the signed difference between the estimated and nominal distance between the spheres. Up to 0.3% of the worst points are rejected.
- Flatness, F, which is the range of residuals from the measured points to a least squares fitted plane, with up to 0.3% of the worst points rejected.

P_F and P_S are measured using one of the spheres at 10 positions within the system's FOV. SD is measured with the ball-bar at 7 positions, while F is determined using the flat in 6 positions. These positions are illustrated in Figure 4.

Parameter	Value
Center distance	198.9612 mm
Distance uncertainty	0.001 mm
Diameter ball 1	24.9989 mm
Diameter ball 2	24.9969 mm
Min. dev. from sphere 1	-0.0013 mm
Max. dev. from sphere 1	$+0.0006$ mm
Min. dev. from sphere 2	-0.0011 mm
Max. dev. from sphere 2	$+0.0020$ mm
Deviation uncertainty	0.0018 mm

Table 2: Specification of the dumbbell used for our experiments.

4. CALIBRATION PARAMETERS

The industry standard models that are essential for calibration of an SL system contain several parameters. Which of these parameters to include in the calibration process is unclear. To solve for

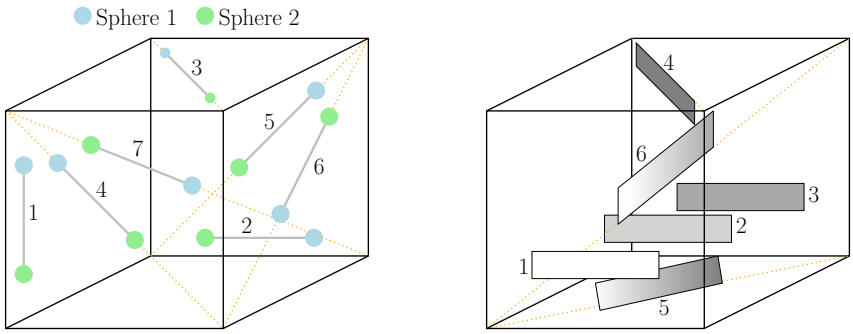


Figure 4: Measurement positions used throughout the paper. The outer frame represents the FOV, as seen from the cameras (Position 1 being closest). Left: ball-bar positions used for sphere distance SD. Right: positions of the flat used for the flatness error metric, F.

Parameter	Value
Minimum deviation from plane	-0.0030 mm
Maximum deviation from plane	$+0.0012$ mm
Deviation uncertainty	0.0018 mm

Table 3: Specification of the flat plane used for our experiments.

the calibration parameters we employ the commonly used method proposed by Zhang (Zhang, 2000). We use the 4 parameter pinhole model with the addition of up to five lens distortion parameters. Hence, the camera is modeled as

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

The use of a non-unit aspect ratio (*i.e.*, $f_x \neq f_y$), makes it possible to model non-square pixels and/or capture compound non-uniformity in the lens. Likewise estimation of the principle point, (c_x, c_y), makes it possible to describe cameras in which the principle ray does not strike the image sensor in it's exact center. With quality components such as ours, we would expect these parameters to be unnecessary. At the same time, the inclusion of these parameters increases the risk of false estimation, numerical instability and non-convergence. In fact, it was shown, that principle point estimation is especially prone to misinterpretation, and that the parameter can often be neglected in cameras of medium to long focal length (Ruiz et al., 2002).

Radial lens distortion is modeled according to

$$x' = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{2}$$
$$y' = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad , \tag{3}$$

where (k_1,k_2,k_3) are the three distortion coefficients. Tangential distortion is modeled

$$x' = x + (2p_1 xy + p_2(r^2 + 2x^2)) \tag{4}$$
$$y' = y + (2p_1(r^2 + 2y^2) + 2p_2 xy) \tag{5}$$

where (p_1, p_2) are the tangential distortion parameters. This five parameter "Brown-Conrady" model is widely accepted (Brown, 1966).

The stereo relationship between cameras is described using three rotations and three translations. Due to weak inter-dependencies, the calibration can be performed individually per camera, followed by stereo calibration. Still, the risk of over-fitting and converging to local minima remains, and therefore higher order distortion parameters are used only when considered relevant. To investigate these factors, we calibrate using 8 different configurations of parameters and evaluate by means of VDI/VDE quality parameters. Each calibration is performed using 81 observations of the calibration board, evenly sampled in the range from -40 to 40 degrees relative to baseline.

Figure 5 shows performance results for the different calibration parameter configurations. The baseline setting generally yields sub-millimeter results. The free aspect ratio ($f_x \neq f_y$) and principal point estimation degrade the performance from "baseline". These results show that in a typical setup, omitting the principle point estimation makes calibration significantly more stable. It can be seen that by enabling the first two distortion coefficients, significant improvement is obtained. This is especially noticeable in the sphere distance metric, SD, being a measure of accuracy. No significant improvement is obtained with additional distortion parameters.

Conclusion Given our setup, only the k_1 and k_2 distortion coefficients are required for accurate calibration. The inclusion of both aspect ratio and principle point estimation makes the calibration procedure unstable, and considerably better results are obtained without them. With their removal, we see consistently low results of P_F, SD and F, while the estimation of sphere sizes (P_S) is biased to positive values. This indicates that one should carefully consider which camera model is used.

5. CALIBRATION OBSERVATIONS

An important question in calibration is in which poses the calibration board needs to be observed. Viewing the calibration board at very shallow angles means higher uncertainty in point localization. In addition, the effect of non-planarity becomes larger. However, it is necessary to observe some degree of foreshortening for focal length estimation (Zhang, 2000).

In this section we attempt to obtain the optimal angular range of observations relative to the baseline. We tested 8 different ranges starting from -5° to 5° relative to baseline and ending in -40° to 40°. For each range, we sample evenly 11 images of the calibration board. For the rotations performed, most foreshortening will be observed around the rotation axis, thus constraining the focal length parameter f_x well. With a fixed aspect ratio, this in turn constraints f_y also.

The results from the experiment can be seen in Figure 6. It is seen that increased foreshortening affects the sphere distance parameter (SD) positively indicating better calibration. In general, the results are quite comparable for all ranges. Comparing to Figure 5 it is also apparent that using 11 observations and 81 observations ranging from -40 to 40 yields similar results.

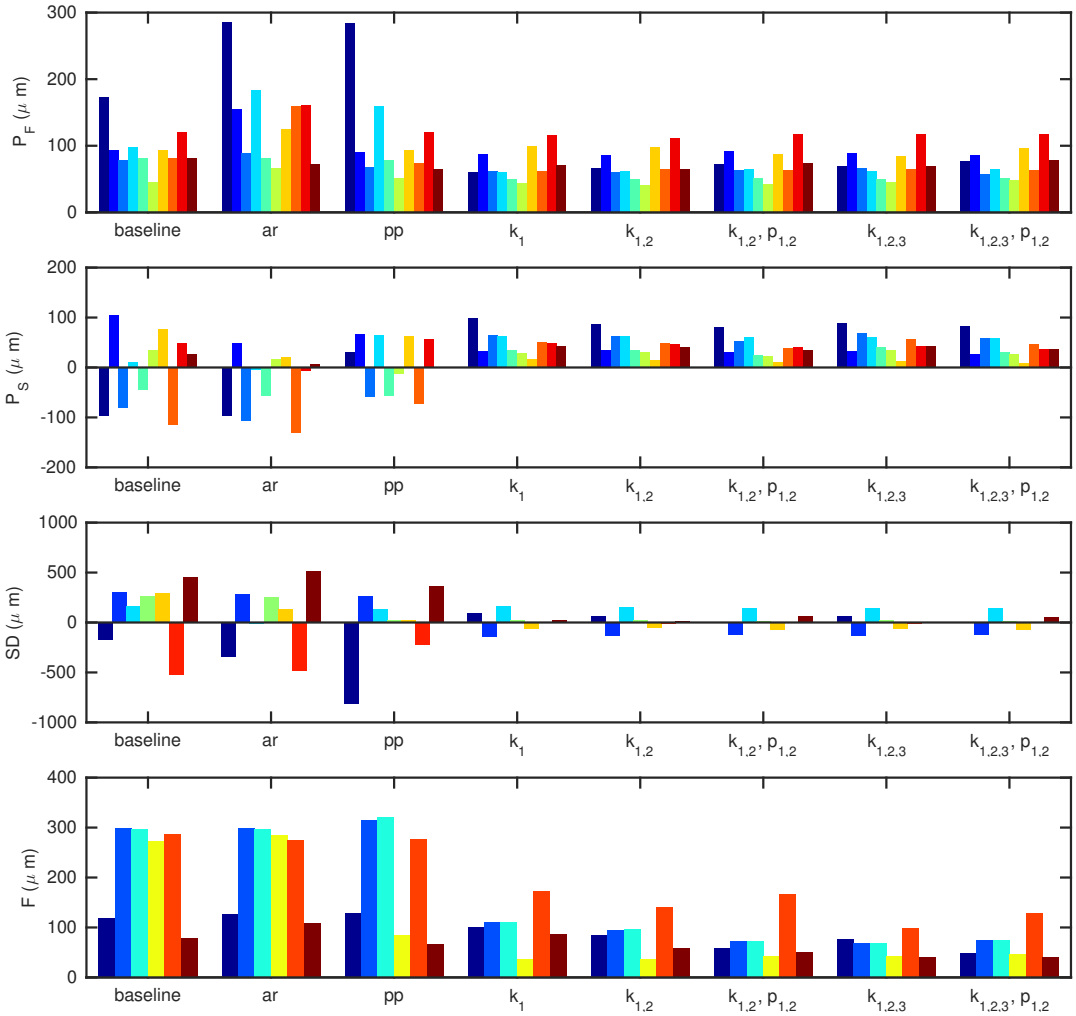


Figure 5: Results obtained with different camera and lens models. Colors represent different positions of the dumbbell or flat artifact according to Fig. 4 (Position 1 being the leftmost bar). Baseline denotes the pinhole model with fixed aspect ratio, fixed principle point and without distortion parameters. ”ar” adds aspect aspect ratio, ”pp” principle point determination. The other groups show results when different combinations of lens parameters are used. From this we see significant improvements when lens distortion parameters are added. The inclusion of both aspect ratio and principle point estimation makes the calibration procedure unstable, and considerably better results are obtained without them.

Conclusion In terms of accuracy, it is slightly beneficial to use a large angular range during calibration. However, even a smaller amount of foreshortening is sufficient to accurately estimate parameters. We opt for the 80° range. Furthermore, the difference between 11 and 81 observations is negligible, thus for the sake of practicality we proceed by using the former.

6. ENCODING STRATEGIES

The encoding strategy of a structured light system determines how correspondences are found, and can be expected to be a major factor in system precision. We identify three main categories of algorithms which are relevant in this setting:

- Fourier methods, prominently phase shifting (PS) methods (Srinivasan et al., 1985).
- Binary boundary based methods (Posdamer and Altschuler, 1982), *e.g*. Binary and Gray coding.
- Line shifting (Gühring, 2000b), which is the same principle underlying triangulation based laser line scanners, with multiple lines sweeping the scene simultaneously.

Phase Shift (PS) based methods encode the scene by means of a series of shifted sinusoidal patterns. The phase is then recovered and matched between cameras (Kühmstedt et al., 2007). The advantage is that the scene can be densely encoded using at least 3 patterns, and more can be naturally added to increase precision. For correct encoding, the projector-camera system should have a close to linear intensity-response. The frequency of sinusoids can also be altered, with higher frequencies yielding higher precision at the cost of phase ambiguities, which then have to be "unwrapped" using additional patterns. Our PS implementation performs 32 steps at the highest frequency sinusoidal pattern (period $19.2\,\mathrm{px}$), and unwraps the resulting phase using two sets of lower frequency patterns (Zumbrunn, 1987). The total number of projected patterns is 64.

Binary boundary based methods, such as the Gray Code method, encode scene points directly by means of binary codes, which are decoded and matched in the cameras. These methods are flexible in the number of patterns, and allow for the natural addition of redundant information, which can then be used to reject outliers. Feature points locations can be estimated with sub-pixel accuracy. Our Gray code implementation encodes uniquely every other column in projector space, and employs patterns and their inverse for added robustness. Boundaries are detected at the intersection of the pattern and its inverse with subpixel accuracy. The total number of patterns is 20.

Line shifting can be performed with a single laser line as the projection source, however with a digital projector, many lines can be projected in parallel. Correspondence points are found at the peak of the stripes. Several methods exist for subpixel peak detection (Trucco et al., 1998). For Gühring's Line Shifting method, we employ Gray codes to partition the encoded space into 256 unique regions. For each of these regions, a single projector line then walks across it in 8 steps, resulting in a total of 28 patterns. The peak of each single line is determined as the first derivative zero crossing using a Gaussian derivative filter of size $5\,\mathrm{px}$.

These classes of encoding strategies have fundamentally different error characteristics. The binary and line shifting methods may be very robust against point outliers, but PS patterns are often less affected by projector and camera blur due to their low-frequency nature.

Parameter	PS	Gray	Line Shift
Number of patterns	64	20	28
Off-focus robustness	Excellent	Good	Moderate
Precision	Excellent	Good	Good
Accuracy	Excellent	Excellent	Excellent
Nr. of points (sphere)	$\sim 20\,\mathrm{k}$	$\sim 7\,\mathrm{k}$	$\sim 10\,\mathrm{k}$
Nr. of points (flat)	$\sim 450\,\mathrm{k}$	$\sim 150\,\mathrm{k}$	$\sim 250\,\mathrm{k}$

Table 4: Interpretation of the algorithm performance.

The SL system is calibrated with the previously determined angular range of 80 degrees and 11 positions. Furthermore, we use the $k_{1,2}$ parameter selection, as previously identified. A comparison of the VDI/VDE quality parameter results for these three encoding strategies is seen in Figure 7. A summary of the results may be seen in Table 4.

The precision of the different strategies can be indirectly estimated from the spherical form parameter P_F. This is because the calibration spheres cover only a small part of the scanning volume, whereas the flat plane occupies a substantial part. Flat plane measurements will thus be more affected by the quality of calibration and lens distortion correction. Both of which directly affect precision and accuracy.

From the results we confirm that the PS method is more tolerant to depth of field limitations, where positions close and far away show no signs of degradation. The PS method also shows superior precision characteristics in the P_F parameter. For the P_S parameter, there is a clear bias present in both the PS and Gray code method, whereas Line Shift appears bias free. Figure 8 illustrates sphere fitting results for the three methods. Here it is seen, that PS and Gray have systematic positive residuals towards the lateral edge in horizontal (encoding) direction. This in turn leads to slight overestimation of sphere diameters.

Conclusion In the quality parameters P_F (Spherical form) and F (Flatness), the PS encoding strategy provides the best results in all artifact positions. The sphere diameter is biased positively in the PS and Gray methods, while it appears unbiased with Line Shift. Overall it appears that the PS method is the best performing method.

7. COMPARISON TO METROLOGY SCANNER

Finally, we have compared our system to a high-end commercial scanner (GOM ATOS III Triple Scan), which has a FOV of $240 \times 320 \times 240\,\mathrm{mm}$, similar to the FOV of our system (see Table 1).

In this experiment we used the PS algorithm, and calibrated as in the preceding experiment. Again, quality measures defined by the VDI/VDE 2634(2) were measured. Results are seen in Figure 9. The results show that our system is more precise and in terms of P_F and exhibits lower variance. However, a bias is present in the P_S whereas the commercial system appears free of such. It is apparent from the sphere spacing term (SD) that the commercial system indicates better accuracy.

Interestingly, the GOM scanner shows significant improvements in the flatness form error metric, F, compared to the sphere form, P_F, where one would expect similar performance (as seen in our system). Reasons for this will only be cause for speculation as the scanning procedure and reconstruction is proprietary, that being said, some form of smoothing favoring planar surfaces might be at work.

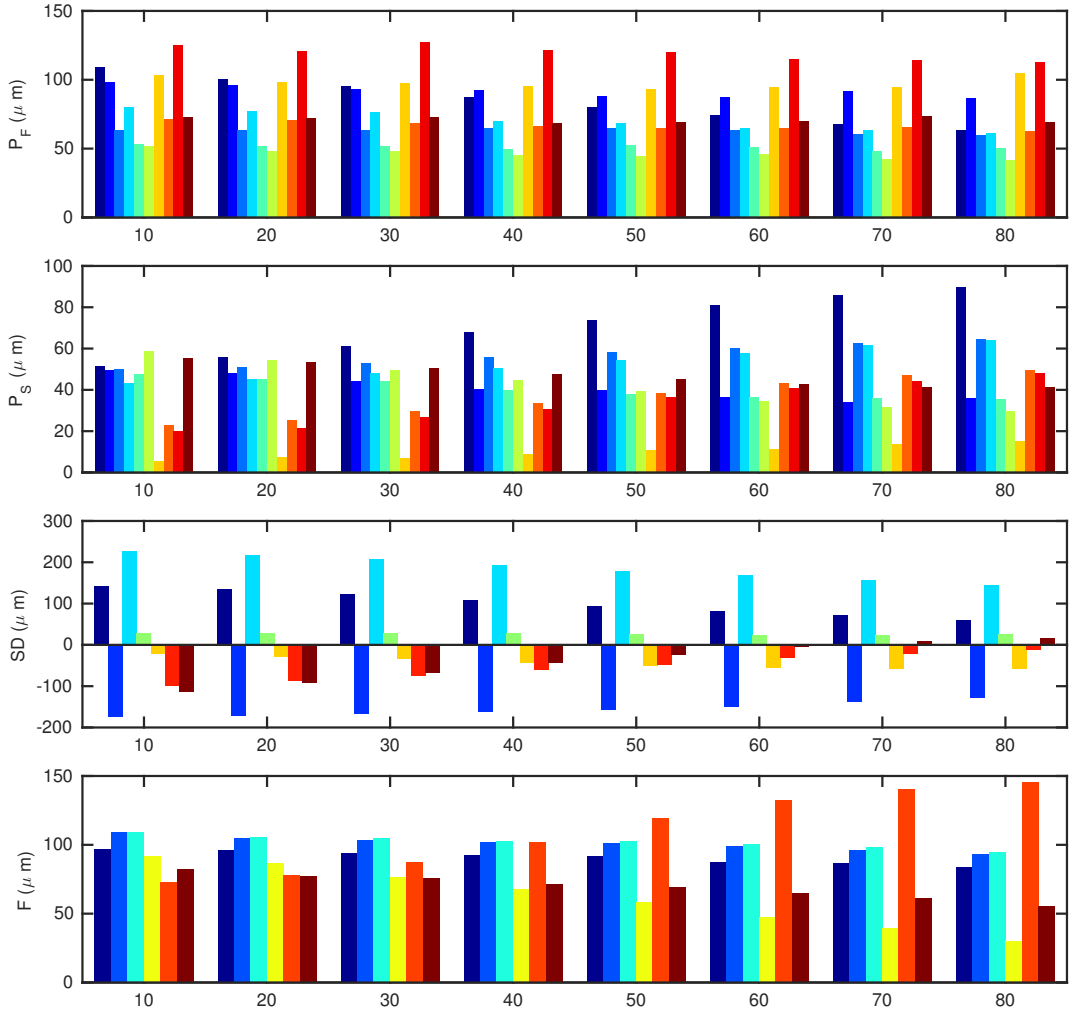


Figure 6: Results obtained with different angular ranges of the calibration plate relative to the camera baseline. Colors represent different positions of the artifacts according to Fig. 4 (Position 1 being the leftmost bar). We see that in terms of accuracy, it is slightly beneficial to use a large angular range during calibration. However, even a smaller amount of foreshortening is sufficient to accurately estimate parameters.

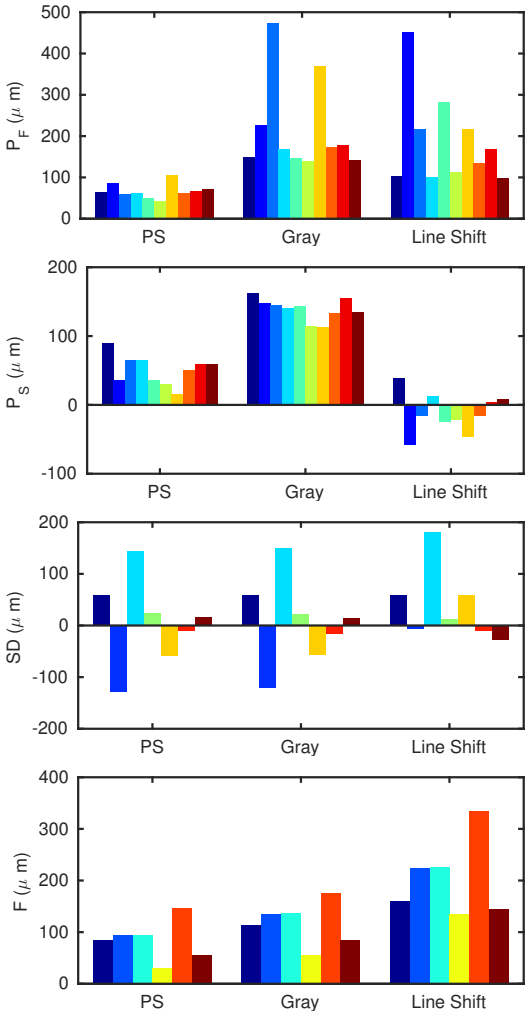


Figure 7: Results obtained with different encoding strategies. Colors represent different positions of the ball-bar or flat artifact according to Fig. 4 (Position 1 being the leftmost bar). Based on this we conclude the PS method to be the overall best performer, yielding significantly higher precision.

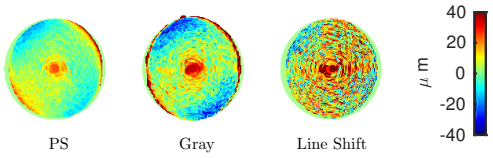


Figure 8: Fitted spheres in position 1 for our three algorithms. In PS and Gray, a bias is observed toward the edges in the horizontal (pattern encoding/epipolar line) direction, while Line Shift does not suffer from this bias. This results in an over estimation of the diameter, P_S, as seen in Figure 7.

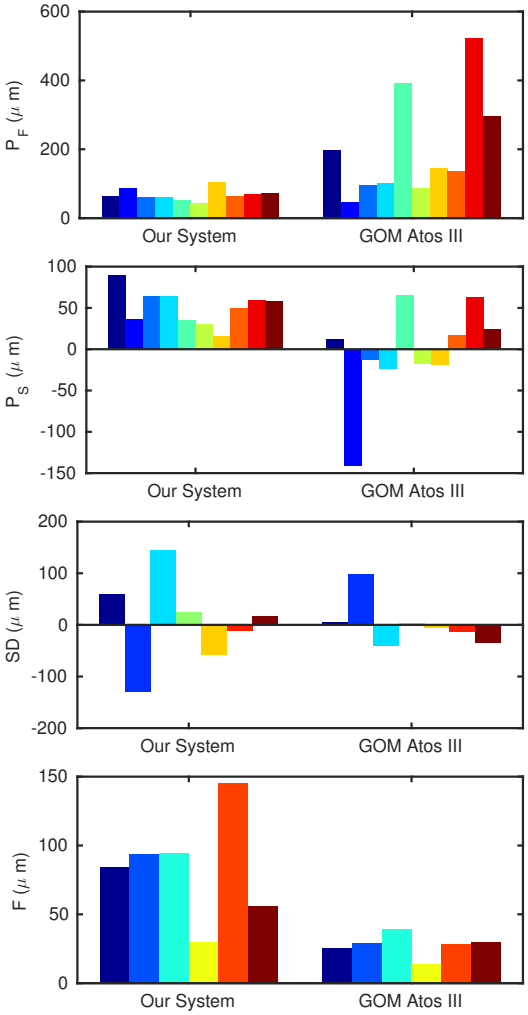


Figure 9: Results obtained with our system compared to results from a commercial metrology grade structured light scanner (GOM Atos III). Colors represent different positions of the dumbbell or flat artifact according to Fig. 4 (Position 1 being the leftmost bar).

Conclusion Our system generally performs better in the precision characteristic, P_F, while the the metrology scanner obtains unbiased sphere diameter results and achieves higher accuracy. Since accuracy is a deterministic noise component, this indicates that a custom calibration method could be advantageous.

8. DISCUSSION AND CONCLUSION

In conclusion, in this paper we have shown through quantitative analysis how the most common parameters within structured light systems affect the overall performance. Our quantitative measure is the accepted VDI/VDE 2634(2) guideline which nicely captures critical parameters in terms of precision and accuracy. We perform a series of experiments on our experimental setup using precision made calibration artifacts. We start by investigating calibration parameters as defined by the most commonly used models and follow by determining the angular foreshortening and the amount of observations required to yield the best results. We proceed by comparing three commonly used algorithms against each other in order to determine the best method. Finally, we compare our setup to a metrology grade commercial scanner, using the previously determined parameters. Our results show that comparable and in some cases better results can be obtained using standard methods and models if care is taken in the parameters choice. We expect these findings to be of help to practitioners wanting to build their own SL systems.

Even though the VDI guideline indirectly captures some of the error sources, such as depth of field, calibration performance and acquisition noise, it is lacking to some extent. The calibration shapes suggested consists of low frequency features, thus a Gaussian filtering operation on the measured point cloud will yield better results for some of the parameters. Although it is stated that all filtering operations must be noted; In many cases these filtering operations are required or inherent in the triangulation algorithms at a pre point cloud level. The use of such filtering will affect the frequency response of the system, where low-pass operations will limit the systems capability of resolving high frequency features. In order to better analyze a systems performance, a frequency analysis must be conducted, indicating if any such smoothing is taking place. Such a frequency response characterization calls for an additional calibration artifact in the form of a high frequency feature. In addition, results from the VDI/VDE only provide quantitative evaluation of the artifacts used. Thus the results cannot be transferred to other less optically ideal materials.

REFERENCES

2634, V., 2000. Optical 3d measuring systems. VDI/VDE guideline.

Beraldin, J.-A. and Gaiani, M., 2005. Evaluating the performance of close-range 3d active vision systems for industrial design applications. In: Electronic Imaging 2005, International Society for Optics and Photonics, pp. 67–77.

Beraldin, J.-A., Blais, F., El-Hakim, S., Cournoyer, L. and Picard, M., 2007. Traceable 3d imaging metrology: Evaluation of 3d digitizing techniques in a dedicated metrology laboratory. In: Prooceedings of the 8th Conference on Optical 3-D Measurement Techniques.

Boehm, J., 2014. Accuracy investigation for structured-light based consumer 3d sensors. Photogrammetrie-Fernerkundung-Geoinformation 2014(2), pp. 117–127.

Brown, D. C., 1966. Decentering distortion of lenses. Photometric Engineering 32(3), pp. 444–462.

Chen, S., Li, Y. F. and Zhang, J., 2008. Vision processing for realtime 3-d data acquisition based on coded structured light. IEEE Transactions on Image Processing, 17(2), pp. 167–176.

Fisher, R. B., Ashbrook, A., Robertson, C. and Werghi, N., 1999. A low-cost range finder using a visually located, structured light source. In: Proc. 3-D Digital Imaging and Modeling, Vol. 1, IEEE, pp. 24–33.

Geng, J., 2011. Structured-light 3d surface imaging: a tutorial. Advances in Optics and Photonics 3(2), pp. 128–160.

Gorthi, S. S. and Rastogi, P., 2010. Fringe projection techniques: whither we are? Optics and Lasers in Engineering 48(IMAC-REVIEW-2009-001), pp. 133–140.

Gühring, J., 2000a. Dense 3d surface acquisition by structured light using off-the-shelf components. In: Photonics West 2001, International Society for Optics and Photonics, pp. 220–231.

Gühring, J., 2000b. Dense 3D surface acquisition by structured light using off-the-shelf components. Photonics West 2001 pp. 220—-231.

Guidi, G., 2013. Metrological characterization of 3d imaging devices. In: SPIE Optical Metrology 2013, International Society for Optics and Photonics, pp. 87910M–87910M.

Heikkila, J. and Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. In: Proc. CVPR 1997, IEEE, pp. 1106–1112.

Kühmstedt, P., Munckelt, C., Heinze, M., Bräuer-Burchardt, C. and Notni, G., 2007. 3d shape measurement with phase correlation based fringe projection. In: Optical Metrology, International Society for Optics and Photonics, pp. 66160B–66160B.

Legarda-Sa, R., Bothe, T., Ju, W. P. et al., 2004. Accurate procedure for the calibration of a structured light system. Optical Engineering 43(2), pp. 464–471.

Li, Z., Shi, Y., Wang, C. and Wang, Y., 2008. Accurate calibration method for a structured light system. Optical Engineering 47(5), pp. 053604–053604.

Lohry, W., Xu, Y. and Zhang, S., 2009. Optimal checkerboard selection for structured light system calibration. In: SPIE Optical Engineering, International Society for Optics and Photonics, pp. 743202–743202.

Luhmann, T., 2011. 3d imaging: how to achieve highest accuracy. In: SPIE Optical Metrology, International Society for Optics and Photonics, pp. 808502–808502.

Posdamer, J. and Altschuler, M., 1982. Surface measurement by space-encoded projected beam systems.

Rocchini, C., Cignoni, P., Montani, C., Pingi, P. and Scopigno, R., 2001. A low cost 3d scanner based on structured light. In: Computer Graphics Forum, Vol. 20number 3, pp. 299–308.

Ruiz, A., López-de Teruel, P. E. and García-Mateos, G., 2002. A note on principal point estimability. In: Proc. Int. Conf. on Pattern Recognition, Vol. 2, IEEE, pp. 304–307.

Salvi, J., Pages, J. and Batlle, J., 2004. Pattern codification strategies in structured light systems. Pattern Recognition 37(4), pp. 827–849.

Srinivasan, V., Liu, H.-C. and Halioua, M., 1985. Automated phase-measuring profilometry: a phase mapping approach. Applied Optics 24(2), pp. 185–188.

Trucco, E., Fisher, R. B., Fitzgibbon, A. W. and Naidu, D., 1998. Calibration, data consistency and model acquisition with laser stripers. Int. Journal of Computer Integrated Manufacturing 11(4), pp. 293–310.

Tsai, M.-J. and Hung, C.-C., 2005. Development of a high-precision surface metrology system using structured light projection. Measurement 38(3), pp. 236–247.

Tsai, R. Y., 1987. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. IEEE Journal of Robotics and Automation 3(4), pp. 323–344.

Zhang, S. and Huang, P. S., 2006. Novel method for structured light system calibration. Optical Engineering 45(8), pp. 083601–083601.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE PAMI 22(11), pp. 1330–1334.

Zumbrunn, R., 1987. Automated fast shape determination of diffuse reflecting objects at close range, by means of structured light and digital phase measurement. In: ISPRS Conference on Fast Processing of Photogrametric Data, pp. 363–379.

Appendix H

Structured Light Scanning of Skin, Muscle and Fat

Technical Report DTU Compute Technical Report-2015-07
Published in DTU Orbit

The technical considers the bias and variance in SL scans of biological tissue, related to subsurface effects as discussed in Ch. 6. This is relevant in various industrial, life-science and medical applications of SL, where subsurface effects bias the scan result. We quantify the amount of bias observed with different scene coding methods, and shows how it can be removed from the scan, resulting in a clear improvement of scan accuracy.

DTU Compute Technical Report-2015-07

Structured Light Scanning of Skin, Muscle and Fat

Jakob Wilm, Sebastian Nesgaard Jensen, and Henrik Aanæs

Department of Applied Mathematics and Computer Science
Technical University of Denmark

Abstract

We investigate the quality of structured light 3D scanning on pig skin, muscle and fat. These particular materials are interesting in a number of industrial and medical use-cases, and somewhat challenging because they exhibit subsurface light scattering. Our goal therefor is to quantify the amount of error that various encoding strategies show, and propose an error correcting model, which can bring down the measurement bias considerably. Samples of raw and unprocessed pig tissue were used with the number of sampled surface points $N_{\text{meat}} = 1.2 \cdot 10^6$, $N_{\text{skin}} = 4.0 \cdot 10^6$ and $N_{\text{fat}} = 2.1 \cdot 10^6$ from 8 different pieces of tissue. With the standard N-step phase shifting method, the bias and RMS errors were found to be 0.45 ± 0.22 mm (muscle), 0.51 ± 0.19 mm (skin) and 0.14 ± 0.16 mm (fat). After applying a linear correction model containing view, light angles and point distances, the bias was almost completely removed on test data, and standard deviations slightly reduced. To our knowledge this is the first quantitative study of the measurement error of structured light with biological tissue.

1 Introduction

Structured light has become a very popular and versatile method for 3D acquisition of complicated scenes. Many applications have emerged in science and in industry, ranging from 3D model acquisition, to motion capture and high performance metrology. This is due to it's versatility from high accuracy to high speed, in some cases reaching real-time performance [20]. Most structured light methods however rely on the assumption of Lambertian (diffusely reflecting) surfaces and only direct illumination. This means that global lighting effects such as intra-reflections, translucency and subsurface scattering disrupts structured light based surface scanning, causing biased, noisy or missing data.

Human and animal tissue is interesting in a number of structured light applications, including head tracking for medical motion correction [16], to robotic handling of meat in industrial settings. Each tissue type exhibits different reflection and scattering properties that can potentially disrupt a structured light scan. This paper documents the degree to which subsurface effects alter the accuracy of structured light scans in three biological tissue types; pork muscle, pork skin and pork fat. We show, that the scan error consists in part of a large deterministic part, which can be removed by means of a simple statistical error model.

2 Structured Light 3D Scanning

The structured light principle uses a calibrated camera-projector pair to identify feature correspondences and triangulate surface points, see figure 1. A large number of encoding strategies exist (see [5, 6, 18] for recent reviews). Most current high-accuracy methods are either based on binary coding schemes such as Gray codes [17] or based on sinusoidal patterns in so-called "phase shifting" [19], with many alterations

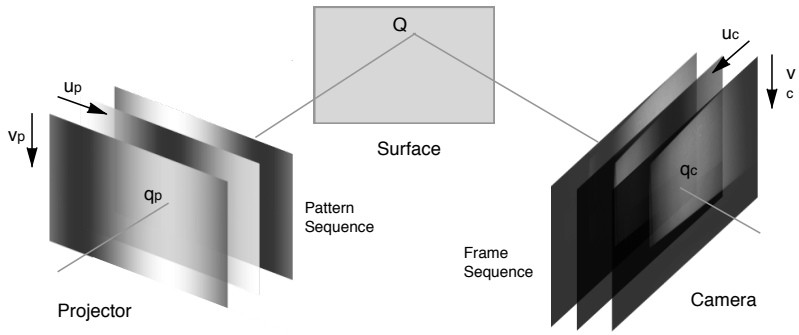


Figure 1: The structured light principle: a number of patterns are projected onto the scene, and images are captured by a camera. Correspondences are determined by different encoding algorithms, and used to triangulate points on the object surface. In this example, 3-step phase shifting patterns are shown.

and improvements to increase their accuracy, robustness or speed. Owing to the epipolar constraint in a calibrated projector-camera pair, encoding needs only to be done along one spatial direction (close to parallel to epipolar lines). With a pattern budget of N patterns, binary methods can uniquely encode 2^N lines. Phase shifting methods use at least $N = 3$ to fully encode the scene at every camera pixel. Additional patterns are used to increase the SNR, most often by projecting multiple phases of a sinusoid, and additional patterns to disambiguate between these phases in "temporal phase unwrapping". Binary coding methods are considered robust against many unwanted illumination effects, but do not make very efficient use of the pattern budget, and also fail once light stripes become too narrow. Phase shifting methods make efficient use of patterns, but provide no general means of identifying encoding errors, possibly leading to noise or outliers.

3 Global lighting effects

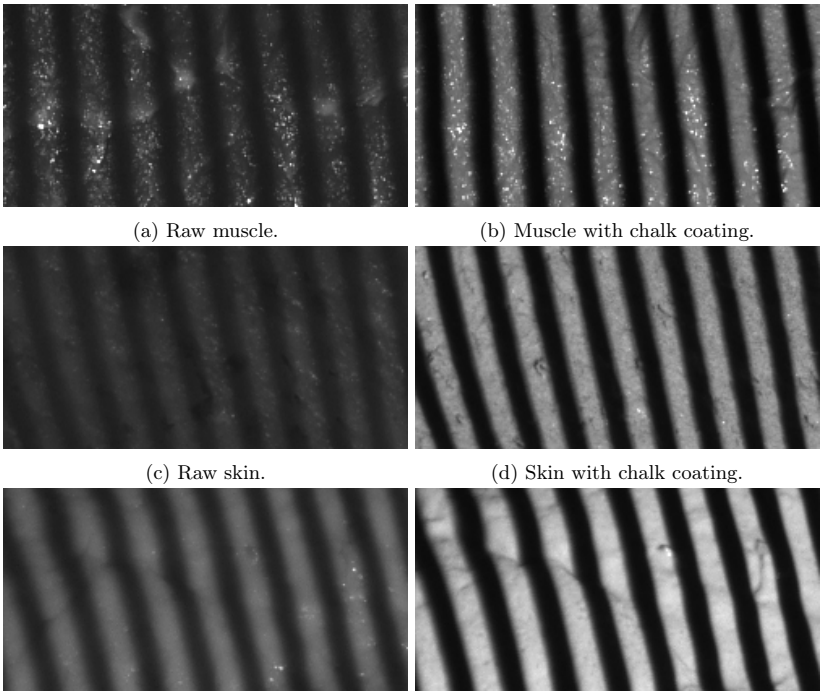
Under ideal circumstances, objects to be scanned with structured light exhibit only direct reflection on the object surface. Also, in order to not introduce under or over-saturation or bias in feature detection, the surface reflectance ideally is high, and Lambertian, i.e. diffusely reflecting into all directions. Under real conditions, most interesting objects exhibit global lighting effects, that is light does not reach the camera after a single scattering event on the object surface, but does so after reflecting from a distant scene point or is scattered below the surface (so-called sub surface scattering).

While some solutions have been proposed to mitigate subsurface scattering, defocus and global illumination effects [1, 2, 7], such effects are still considered harmful, and many object surfaces need to be treated by spraying them with "dulling spray" in order to make structured light possible. An even more effective way of reducing global lighting effects is by using chalk spray, which makes surfaces highly reflective and diffuse. figure 2 shows sample images of structured light patterns projected onto our samples before and after treatment with chalk-spray.

In particular the muscle samples show large degrees of subsurface scattering, which blurs and distorts the structured light patterns.

4 Related works

The issue of global lighting effects in the context of structured light has been recognised by many authors, e.g. in the acquisition of a human face reflectance field [4]. In order to reduce these in structured light scans, polarisation has been exploited in [1]. Some recent attempts have been to design structured light encoding strategies such that they are less susceptible to global lighting effects. The underlying observation is, that with high-frequent patterns, global lighting effects can be considered constant, and independent of the actual pattern. This allows for efficient separation of the observed light intensities



(a) Raw muscle. (b) Muscle with chalk coating.

(c) Raw skin. (d) Skin with chalk coating.

(e) Raw fat. (f) Fat with chalk coating.

Figure 2: Fine grained binary structured light pattern projected onto various types of tissues. The effect of subsurface scattering is clearly seen the pattern becomes blurred without chalk coating.

into direct and global light [15]. In modulated phase shifting [2], structured light patterns are modulated by means of carrier patterns, such that they become high-frequent in both spatial dimensions, thereby improving their separation power. Micro Phase Shifting [9] makes use of sinusoidal patterns in a narrow high-frequency band, promising robustness to global lighting effects and stable phase unwrapping with an optimal number of patterns. A newer approach is "unstructured light" [3], in which the pattern frequency can be high in both dimensions, however the number of patterns is not ideal, and the matching procedure rather cumbersome. For binary boundary methods, exclusively high or low-frequency pattern schemes can be considered robust against different global illumination effects [8].

An approach to compensate for the measurement error in isotropic semi-transparent material caused by subsurface scattering was presented in [13]. Similarly to our approach, this work empirically determines the measurement error and explains it by means of a single variable (the projected light angle), albeit only with a single verification object and structured light method. In [12], a Monte-Carlo simulation of the measurement situation was presented, which gives some insight into the error forming process.

In [10], an analytical derivation of the measurement error is given for the phase shifting methods. This error model predicts the error to decrease with increased spatial frequency of the pattern, agreeing with the theory of direct-global illumination. The model does not however take into account the loss of amplitude at higher frequency patterns, which increases noise in the measurement data.

Computer simulations of structured light scans were performed by [14] to benchmark encoding methods with respect to various parameters, and were found to have similar robustness to subsurface effects.

5 Experiments

A single camera-projector system as shown in figure 1 was used for our experiments. Surface scans were obtained from three separate tissue classes: pork muscle, pork skin and pork fat. All samples were unprocessed and not heat-treated. For each class, 8 distinct tissue samples were measured, once in their raw form, and once with chalk-spray applied. These samples were distinct placed independently in the scan volume and spanned all possible view and light angles, and also varied in their distance to the projector from approximately 200 mm to 400 mm. See figure 4 for an illustration of these parameters. The latter measurements served as a reference for the evaluation of measurement error. While we cannot assume the chalk-prepared surfaces to be noise-free, we consider them the gold-standard as they provide very clear contrast, and virtually no subsurface effects are visible (see figure 2).

In order to verify that this procedure does not alter surface geometry, we applied two separate layers of chalk to a sample object, and compared the scan result after each layer. The mean signed distance was 0.037 mm, indicating that chalk spraying the surfaces does not bias the result.

We analyse four different structured light methods, and use 12 patterns in each:

- Binary Gray coding [17]: one completely lit and one completely dark image were used to define the binary threshold individually in each camera pixel. The remaining patterns were used to encode $2^{10} = 1024$ individual lines on the object surface.
- N-step phase shifting was used with 9 shifts of a high-frequency sinusoid of frequency $1/76\,\mathrm{px}^{-1}$, corresponding to approximately 1/10 mm on the object surface. Three additional patterns were used for phase-unwrapping [11].
- Micro phase shifting [9] using the frequency vector

  $$[1/75.02, 1/70.00, 1/71.32, 1/72.47, 1/73.72, 1/76.23, 1/77.35, 1/78.40, 1/79.22, 1/80.00]\,\mathrm{px}^{-1}$$

  These frequencies corresponds to a spatial frequency on the object surface of approximately 1/10 mm. Slightly different from [9], the specific values were determined using a derivative free non-linear pattern search.
- Modulated phase shifting [2] with three shifts of a sinusoid of frequency $1/76\,\mathrm{px}^{-1}$ (1/10 mm on the object surface). Each of these sinusoids was modulated in the orthogonal direction using a sinusoidal carrier with the same frequency. Three additional patterns were used for phase-unwrapping.

Figure 3 shows the pattern sequences used in our experiments.

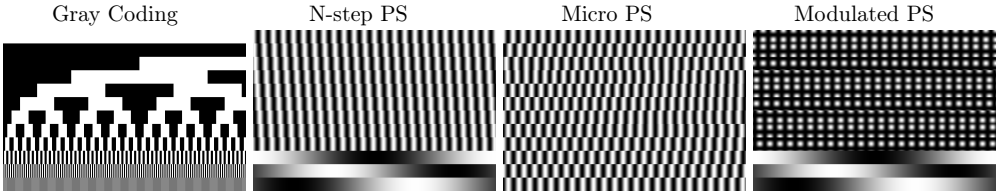


Figure 3: Structured light patterns used in our experiments. In each case, 12 patterns were used.

While binary Gray coding and N-step phase shifting can be considered well-established "reference" methods, the other two phase shifting based methods are state-of-the-art and specifically designed to mitigate subsurface effects. For every sample, we defined a binary mask within which all possible surface points were reconstructed. This ensured that the exact same surface region was used in the evaluation of each method.

Analysis consisted in determining the error in each surface point, by determining its' signed distance to the corresponding point in the chalk-sprayed reference. For Gray-code measurement, we determine the point-to-surface distance as given in the following,

$$s_j = \min_i \left[\boldsymbol{n_i} \cdot (\boldsymbol{p_j} - \boldsymbol{p_i}) \right], \tag{1}$$

where,

$\boldsymbol{p_j}$ is the jth point in the raw point cloud,
$\boldsymbol{p_i}$ is the ith point in the reference point cloud,
$\boldsymbol{n_i}$ is the ith normal in the reference point cloud,
s_j is the signed error at jth data point.

So for each a point we find its' nearest neighbour in the reference cloud in terms of normal distances. This is done as the reconstructed points are not defined in the regular camera pixel grid. As this is the case with the other methods we can simply match points using their position in the pixel grid.

5.1 Error Model

Our principle assumption is that the error is composed of a deterministic part, which once determined can be subtracted from future scans, in order to improve the accuracy. Previous work gives some hints as to which parameters to include [10,13]. Considering the scan setup, as shown in figure 4, we include three variables in our error model: the view angle (relative to the surface normal) θ_{view}, the light angle θ_{light}, and the distance from projector to object, d. We have tried to include many other variables, including reflected light to view angle and coding direction normal vector angles. These variables are inspired by the analytical error model of [10], but did not explain sufficient variance to include them in our model.

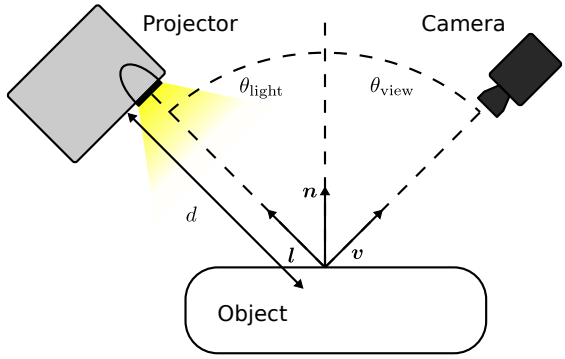


Figure 4: The structured light scanning setup with the parameters of our error model. The surface normal is $\boldsymbol{n}$, view direction $\boldsymbol{v}$, light direction $\boldsymbol{l}$ and the projector-surface distance is d.

We construct the following linear model:

$$y = \beta_0 + \beta_1 \cdot \theta_{\text{view}} + \beta_2 \cdot \theta_{\text{light}} + \beta_3 \cdot d \quad ,$$

for each of the tissue types, and fit it to a training set consisting of 90% of all observations. A test set composed of the remaining 10% was then corrected using the predetermined model.

6 Results

The parameters obtained after fitting the error model to our data are seen in tables 1, 2, and 3. These tables also show the mean signed distance before applying the model, $\epsilon_{\text{training}}$. After applying the linear model, the mean of residuals naturally is zero. We also show the standard deviation of errors before correction using the model, σ_{training}, and after, $\sigma_{\text{corrected}}$.

Figures 5, 6 and 7 show the signed error on a single sample visually before and after applying the correction model. We see qualitatively in the signed errors of Figures 5 – 7, that a clear positive bias is present in all scans (the raw surface is observed further away from the camera). This effect is especially pronounced in the muscle and skin tissue classes and with the N-step phase shifting method (this is a general observation in all muscle and skin samples). Tables 1 and 2 show that the observed bias is

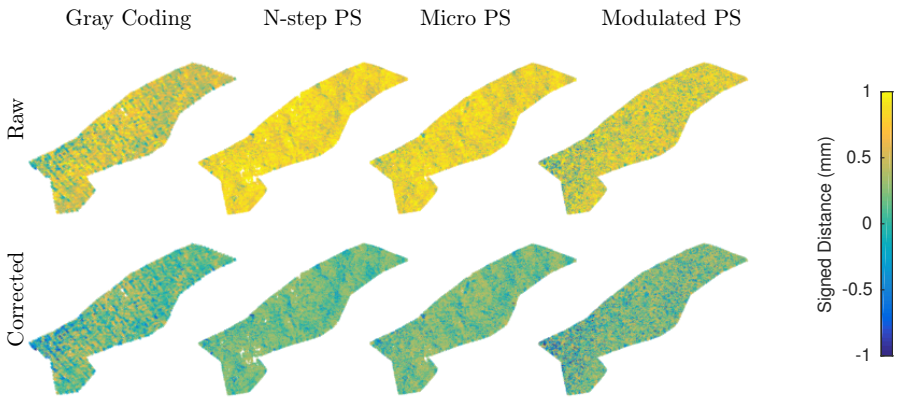


Figure 5: Signed distance (sd) between scan and reference on a single sample of muscle. Top row: before applying the linear correction model. Bottom row: after correction.

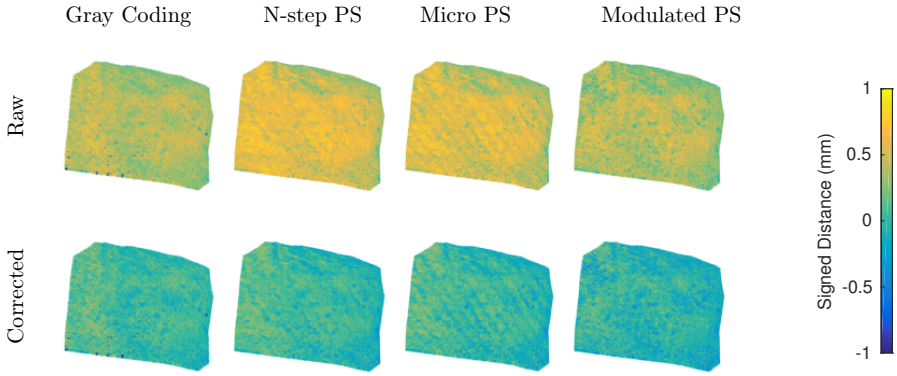


Figure 6: Signed distance (sd) between scan and reference on a single sample of fat. Top row: before applying the linear correction model. Bottom row: after correction.

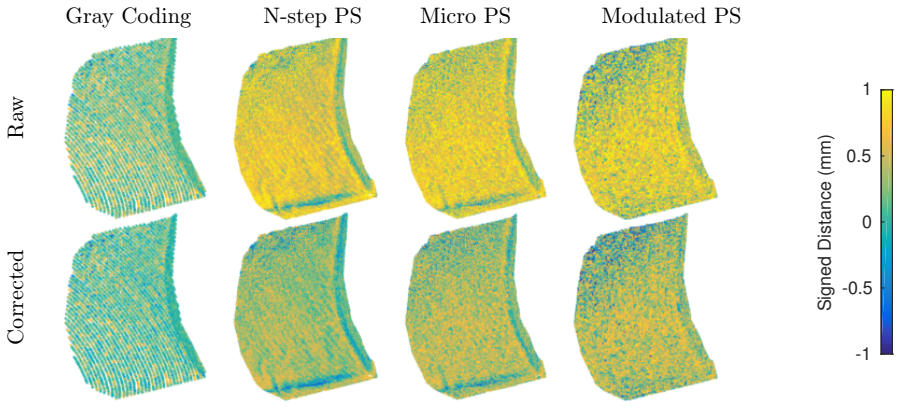


Figure 7: Signed distance (sd) between scan and reference on a single sample of skin. Top row: before applying the linear correction model. Bottom row: after correction.

	Gray Coding	N-step PS	Micro PS	Modulated PS
Intercept	0.25	0.55	0.45	0.39
θ_{view}	-0.064	-0.21	-0.15	-0.0032
θ_{light}	0.03	0.053	0.025	-0.048
d	0.00023	3.9×10^{-5}	5.1×10^{-5}	-7.1×10^{-5}
$\epsilon_{\text{training}}$	0.32	0.45	0.38	0.33
σ_{training}	0.27	0.22	0.23	0.27
$\epsilon_{\text{corrected}}$	-0.00097	-0.00094	0.0006	-0.0014
$\sigma_{\text{corrected}}$	0.27	0.22	0.23	0.26

Table 1: Estimated parameters for the linear correction model for muscle and the mean bias ϵ and standard deviation σ of the training set and the corrected test data set. The number of training samples was $N_{\text{training}} = 1\,122\,534$ and the number of test samples $N_{\text{training}} = 124\,726$.

	Gray Coding	N-step PS	Micro PS	Modulated PS
Intercept	0.038	0.89	1.00	0.76
θ_{view}	-0.0026	-0.15	-0.15	-0.06
θ_{light}	-0.24	-0.14	-0.12	-0.14
d	0.0013	-0.00062	-0.0011	-0.00068
$\epsilon_{\text{training}}$	0.35	0.51	0.48	0.41
σ_{training}	0.19	0.19	0.2	0.22
$\epsilon_{\text{corrected}}$	0.00039	-3.7×10^{-5}	-0.00045	0.00078
$\sigma_{\text{corrected}}$	0.19	0.18	0.19	0.22

Table 2: Estimated parameters for the linear correction model for skin and the mean bias ϵ and standard deviation σ of the training set and the corrected test data set. The number of training samples was $N_{\text{training}} = 3\,603\,495$ and the number of test samples $N_{\text{training}} = 400\,388$.

	Gray Coding	N-step PS	Micro PS	Modulated PS
Intercept	0.082	0.041	0.04	0.07
θ_{view}	-0.083	-0.19	-0.12	-0.081
θ_{light}	-0.019	0.092	0.041	0.026
d	0.00018	0.00037	0.00028	0.00014
$\epsilon_{\text{training}}$	0.087	0.14	0.11	0.097
σ_{training}	0.25	0.16	0.16	0.17
$\epsilon_{\text{corrected}}$	0.00075	0.00037	5.1×10^{-5}	-0.00046
$\sigma_{\text{corrected}}$	0.24	0.16	0.16	0.17

Table 3: Estimated parameters for the linear correction model for fat and the mean bias ϵ and standard deviation σ of the training set and the corrected test data set. The number of training samples was $N_{\text{training}} = 1\,868\,574$ and the number of test samples $N_{\text{training}} = 207\,619$.

around 0.6 mm for phase shifting based methods in muscle and skin, while it is only 0.32 mm for Gray coding. The same superiority of Gray coding is observed in the fat tissue class, albeit with generally much lower numbers. This is presumably due to the fact that fat appears more optically dense (see figure 2).

While some spatial structure is visible in the bias seen on Figures 5 – 7, the included variables, θ_{view}, θ_{light}, d, did only explain a small part of the variance. Hence, $\epsilon_{\text{corrected}}$ is reduced significantly from $\epsilon_{\text{training}}$, while that is not the case with $\sigma_{\text{corrected}}$ and σ_{training}.

7 Conclusion

In this work, we performed structured light scans of biological tissue, pork muscle, pork skin and pork fat, with a selection of reference and modern structured light methods. We used a linear error model to describe measurement error, and found it to be composed of a significant bias (≈ 0.5 mm for muscle, ≈ 0.3 mm for skin, and ≈ 0.1 mm for fat), which can be corrected easily. This methodology allowed us to improve the accuracy of structured light scans, given the specific tissue type. The error correction

method proposed is applicable to any type of homogenous material. While the linear parameters need to be determined beforehand, and vary widely between materials, many applications do involve only homogenous material with uniform optical properties.

As such, we consider our finding interesting information on the accuracy that can be obtained with structured light in general, and our methodology to apply in many common scenarios.

References

[1] Tongbo Chen, Hendrik P A Lensch, Christian Fuchs, and Hans Peter Seidel. Polarization and phase-shifting for 3D scanning of translucent objects. *Proc. IEEE CVPR*, 2007.

[2] Tongbo Chen, Hans-Peter Seidel, and Hendrik P.A. Lensch. Modulated phase-shifting for 3D scanning. *Proc. IEEE CVPR*, pages 1–8, 2008.

[3] Vincent Couture, Nicolas Martin, and Sébastien Roy. Unstructured light scanning robust to indirect illumination and depth discontinuities. *Int. Journal on Computer Vision*, 108(3):204–221, 2014.

[4] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. *Proc. SIGGRAPH*, pages 145–156, 2000.

[5] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 160(2):128–160, 2011.

[6] Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: Whither we are? *Optics and Lasers in Engineering*, 48(2):133–140, 2010.

[7] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. Structured light 3D scanning in the presence of global illumination. *Proc. IEEE CVPR*, pages 713–720, jun 2011.

[8] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. A Practical Approach to 3D Scanning in the Presence of Interreflections, Subsurface Scattering and Defocus. *Int. Journal on Computer Vision*, 102(1-3):33–55, aug 2012.

[9] Mohit Gupta and Shree K Nayar. Micro Phase Shifting. *Proc. IEEE CVPR*, pages 813–820, 2012.

[10] Michael Holroyd and Jason Lawrence. An Analysis of Using High-Frequency Sinusoidal Illumination to Measure the 3D Shape of Translucent Objects. *Proc. IEEE CVPR*, pages 2985–2991, 2011.

[11] J M Huntley and H Saldner. Temporal phase-unwrapping algorithm for automated interferogram analysis. *Applied Optics*, 32(17):3047–3052, 1993.

[12] Peter Lutzke, Stefan Heist, Peter Kühmstedt, Richard Kowarschik, and Gunther Notni. Monte Carlo simulation of three-dimensional measurements of translucent objects. *Optical Engineering*, 54(8):084111, 2015.

[13] Peter Lutzke, Peter Kühmstedt, and Gunther Notni. Measuring error compensation on three-dimensional scans of translucent objects. *Optical Engineering*, 50(6):063601, 2011.

[14] Esdras Medeiros, Harish Doraiswamy, Matthew Berger, and Claudio T Silva. Using Physically Based Rendering to Benchmark Structured Light Scanners. *Pacific Graphics*, 33(7), 2014.

[15] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. on Graphics*, 25(3):935, 2006.

[16] Oline Vinter Olesen, Rasmus R. Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen. Motion tracking for medical imaging: a nonvisible structured light tracking approach. *IEEE Trans. Medical Imaging*, 31(1):79–87, jan 2012.

[17] J.L Posdamer and M.D Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18:1–17, 1982.

[18] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 2010.

[19] V Srinivasan, H C Liu, and M Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Applied Optics*, 24(2):185, 1985.

[20] Jakob Wilm, Oline V. Olesen, and Rasmus Larsen. SLStudio : Open-Source Framework for Real-Time Structured Light. *Proc. IPTA 2014*, 2014.